(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 January 2001 (25.01.2001)

PCT

(10) International Publication Number
WO 01/05998 A1

(51) International Patent Classification⁷: C12P 21/06, 21/02, 21/04, A61K 38/00, 38/16

(21) International Application Number: PCT/US00/19198

(22) International Filing Date: 14 July 2000 (14.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/144,088 16 July 1999 (16.07.1999) US

(71) Applicant *(for all designated States except US)*: **HUMAN GENOME SCIENCES, INC.** [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DUAN, Roxanne, D.** [US/US]; 5515 Northfield Road, Bethesda, MD 20817 (US). **RUBEN, Steven, M.** [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US).

(74) Agents: **KENNY, Joseph, J. et al.**; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/05998 A1

(54) Title: FOLLISTATIN-3

(57) Abstract: The present invention relates to a novel follistatin-3 protein which is a member of the family of inhibin-related proteins. In particular, isolated nucleic acid molecules are provided encoding the human follistatin-3 protein. Follistatin-3 polypeptides are also provided as are vectors, host cells and recombinant methods for producing the same. The invention further relates to screening methods for identifying agonists and antagonists of follistatin-3 activity. Also provided are diagnostic methods for detecting reproductive system-related disorders and disorders of the regulation of cell growth and differentiation and therapeutic methods for treating reproductive system-related disorders and disorders of the regulation of cell growth and differentiation.

Follistatin-3

Field of the Invention

The present invention relates to a novel human gene encoding a polypeptide which is a member of the family of inhibin-related proteins. More specifically, isolated nucleic acid molecules are provided encoding a human polypeptide named follistatin-3. Follistatin-3 polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. Also provided are diagnostic methods for detecting disorders related to the reproductive system, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying agonists and antagonists of follistatin-3 activity.

Background of the Invention

The family of inhibin-related proteins currently consists of at least four groups of members: inhibins, activins, and two splice variants of follistatin-1 (315 and 288 amino acids). Inhibins and activins are members of the transforming growth factor (TGF)-beta superfamily and function with opposing actions in a variety of capacities in paracrine and autocrine regulation of both reproductive and nonreproductive organs including the liver, kidney, adrenal glands, bone marrow, placenta, anterior pituitary, and brain (Ying, S. Y., *et al., Proc. Soc. Exp. Biol. Med.* **214:**114-122 (1997); Mather, J. P., *et al., Proc. Soc. Exp. Biol. Med.* **215:**209-222 (1997)). Although the follistatins are not closely related to the TGF-beta family, they still play a major role in the follical stimulating hormone (FSH) synthetic pathway by increasing estradiol production and by functioning directly as high affinity activin-binding proteins. Inhibins, activins, and follistatin-1 were all initially identified as regulators of pituitary FSH secretion, but have more recently been further characterized to function as growth factors, embryo modulators, and immune factors (Petraglia, F. Placenta 18:3-8 (1997)). In addition, each of these factors is involved with the regulation of gonadotropin biosynthesis and secretion, ovarian and placental steroidogenesis, and oocyte and spermatogonial maturation (Halvorson, L. M. and DeCherney, A. H. *Fertil. Steril.* **65:**459-469 (1996)).

FSH is a vital component of the regulatory cascade governing development of human oocytes. Primary oocytes in newborns are arrested in the prophase stage of Meiosis I and are surrounded by a 1-2 cell thick layer of follicle cells constituting a structure termed the primordial follicle. In concert with other factors, stimulation of the primordial follicle with FSH initiates its progression to the more complex structures designated the developing and antral follicles (Ueno, N., *et al., Proc. Natl. Acad. Sci. USA* **84**:8282-8286 (1987); Robertson, D. M., *et al., Biochem. Biophys. Res. Comm.* **149**:744-749 (1987)). The antral follicle consists of an enlarged oocyte surrounded by an increased number of follicle cells, a zona pellucida, cortical granules, and a fluid-filled cavity termed the antrum. It is in this state that thousands of developing oocytes are maintained until puberty. Each month following this point, a surge in the local concentration of several additional hormones and other factors, primarily leuteinizing hormone (LH), stimulates accelerates the growth of roughly 15-20 of the developing follicles in the ovary. Only one of these structures will ultimately complete the developmental progression of its enclosed oocyte to the metaphase stage of Meiosis II. The single stimulated follicle will then continue to enlarge until it bursts at the surface of the ovary and releases the oocyte, still surrounded with a coating of follicle cells, for potential fertilization (Bornslaeger, E. A., *et al., Dev. Biol.* **114**:453-462 (1986); Masui, Y. and Clarke, H. J. *Int. Rev. Cytol.* **57**:185-282 (1979); Richards, J. S. *Recent Prog. Horm. Res.* **35**:343-373 (1979)).

Follistatin also plays a central role in the above-described process of follicle development. Follistatin binds stoichiometrically to activins and, as a result, inhibits the activin-induced augmentation of FSH-release from cultured pituitary cells (Kogawa, K., *et al., Endocrinology* **128**:1434-1440 (1991)). Further evidencing a feedback mechanism, cultured granulosa cells produce and secrete follistatin in response to treatment with FSH (Saito, S., *et al., Biochem. Biophys. Res. Comm.* **176**:413-422 (1991); Klein, R., *et al., Endocrinology* **128**:1048-1056 (1991)). Furthermore, it has been determined by synthesizing the results of a number of studies, that follistatin, activin, FSH, LH, and other factors function in concert in a variety of interrelated mechanisms to regulate many developmental processes, including the development of follicles. For example, in the presence of FSH, activin can augment both LH receptor

expression and progesterone production by rat granulosa cells (Sugino, H., *et al.*, *Biochem. Biophys. Res. Comm.* **153**:281-288 (1988)). In addition, activin can significantly enhance the ability of granulosa cells to express FSH receptor and produce inhibin even in the absence of FSH (Nakamura, T., *et al.*, *Biochim. Biophys. Acta* **1135**:103-109 (1992); Sugino, H., *et al.*, *supra*; Hasegawa, Y., *et al.*, *Biochem. Biophys. Res. Comm.* **156**:668-674 (1988)). These and other studies provide support for the idea that follistatin and activin play important roles in the regulation of granulosa cellular differentiation.

In addition to the many well-characterized effects which follistatin, activin, and inhibin elicit on the regulation of various developmental processes in the reproductive system, a large number of studies have more recently begun to define regulatory roles for these molecules in a variety of other tissues and systems. For example, during early embryonic development in *Xenopus laevis*, the action of activin A in developing targets of ciliary ganglion neurons is regulated by localized expression of follistatin (Hemmati-Brivanlou, A. and Melton, D. A. *Nature* **359**:609-614 (1992); Hemmati-Brivanlou, A. and Melton, D. A. *Cell* **77**:273-281 (1994)). In addition, overexpression of follistatin leads to induction of neural tissue (Hemmati-Brivanlou, A., *et al.*, *Cell* **77**:283-295 (1994)). In the mouse, follistatin mRNA is first detected on embryonic day 5.5 in the deciduum, and, subsequently, in the developing hindbrain, somites, vibrissae, teeth, epidermis, and muscle (van den Eijnden-van Raaij, A. J. M., *et al.*, *Dev. Biol.* **154**:356-365 (1992); Albano, R. M., *et al.*, *Development* **120**:803-813 (1994); Feijen, A., *et al.*, *Development* **120**:3621-3637 (1994)). Evidence of the relative importance of such a varied expression of follistatin is provided by Matzuk and colleagues (*Nature* **374**:360-363 (1995)) who demonstrate that follistatin-deficient mice are retarded in their growth, have decreased mass of the diaphragm and intercostal muscles, shiny taut skin, skeletal defects of the hard palate and the thirteenth pair of ribs, their whisker and tooth development is abnormal, they fail to breathe, and die within hours of birth. Since the defects in mice deficient in follistatin are far more widespread than in mice deficient in activin, Matzuk and coworkers (*supra*) suggest that follistatin may modulate the cell growth and differentiation regulatory actions of additional members of the TGF-beta superfamily.

Thus, there is a need for polypeptides that function as regulators of reproductive development, embryonic development, and cell growth and differentiation since disturbances of such regulation may be involved in disorders relating to reproduction and the regulation of cell growth and differentiation. Therefore, there is a need for identification and characterization of such human polypeptides which can play a role in detecting, preventing, ameliorating or correcting such disorders.

Summary of the Invention

The present invention provides isolated nucleic acid molecules comprising a polynucleotide encoding at least a portion of the follistatin-3 polypeptide having the complete amino acid sequence shown in SEQ ID NO:2 or the complete amino acid sequence encoded by the cDNA clone deposited as plasmid DNA as ATCC Deposit Number 209199 on August 8, 1997. The nucleotide sequence determined by sequencing the deposited follistatin-3 clone, which is shown in Figures 1A, 1B, and 1C (SEQ ID NO:1), contains an open reading frame encoding a complete polypeptide of 263 amino acid residues, including an initiation codon encoding an N-terminal methionine at nucleotide positions 19-21, and a predicted molecular weight of about 27.7 kDa. Nucleic acid molecules of the invention include those encoding the complete amino acid sequence excepting the N-terminal methionine shown in SEQ ID NO:2, or the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone in ATCC Deposit Number 209199, which molecules also can encode additional amino acids fused to the N-terminus of the follistatin-3 amino acid sequence.

The encoded polypeptide has a predicted leader sequence of 26 amino acids underlined in Figure 1A; and the amino acid sequence of the predicted mature follistatin-3 protein is also shown in Figure 1A, as amino acid residues 27-263 and as residues 1-237 in SEQ ID NO:2.

Thus, one aspect of the invention provides an isolated nucleic acid molecule comprising a polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding the follistatin-3 polypeptide having the complete amino acid sequence in SEQ ID NO:2 (i.e., positions -26 to 237 of SEQ ID NO:2); (b) a nucleotide sequence encoding the follistatin-3 polypeptide having the

complete amino acid sequence in SEQ ID NO:2 excepting the N-terminal methionine (i.e., positions -25 to 237 of SEQ ID NO:2); (c) a nucleotide sequence encoding the predicted mature follistatin-3 polypeptide having the amino acid sequence at positions 1 to 237 in SEQ ID NO:2; (d) a nucleotide sequence encoding the follistatin-3 polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199; (e) a nucleotide sequence encoding the follistatin-3 polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209199; (f) a nucleotide sequence encoding the mature follistatin-3 polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199; and (g) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), (d), (e) or (f) above.

Further embodiments of the invention include isolated nucleic acid molecules that comprise a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical, to any of the nucleotide sequences in (a), (b), (c), (d), (e), (f) or (g), above, or a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide in (a), (b), (c), (d), (e), (f) or (g), above. This polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues.

An additional nucleic acid embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a follistatin-3 polypeptide having an amino acid sequence in (a), (b), (c), (d), (e) or (f), above. A further embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of a follistatin-3 polypeptide having an amino acid sequence which contains at least one amino acid substitution, but not more than 50 amino acid substitutions, even more preferably, not more than 40 amino acid substitutions, still more preferably, not more than 30 amino acid substitutions, and still even more preferably, not more than 20 amino acid substitutions. Of course, in order of ever-increasing preference, it is highly preferable for a polynucleotide which encodes the

amino acid sequence of a follistatin-3 polypeptide to have an amino acid sequence which contains not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid substitutions. Conservative substitutions are preferable.

The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, and to host cells containing the recombinant vectors, as well as to methods of making such vectors and host cells and for using them for production of follistatin-3 polypeptides or peptides by recombinant techniques.

In accordance with a further aspect of the present invention, there is provided a process for producing such polypeptide by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a follistatin-3 nucleic acid sequence, under conditions promoting expression of said protein and subsequent recovery of said protein.

The invention further provides an isolated follistatin-3 polypeptide comprising an amino acid sequence selected from the group consisting of: (a) the amino acid sequence of the full-length follistatin-3 polypeptide having the complete amino acid sequence shown in SEQ ID NO:2 (i.e., positions -26 to 237 of SEQ ID NO:2); (b) the amino acid sequence of the full-length follistatin-3 polypeptide having the complete amino acid sequence shown in SEQ ID NO:2 excepting the N-terminal methionine (i.e., positions -25 to 237 of SEQ ID NO:2); (c) the amino acid sequence of the predicted mature follistatin-3 polypeptide having the amino acid sequence at positions 1 to 237 in SEQ ID NO:2; (d) the amino acid sequence of the full-length follistatin-3 polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199; (e) the amino acid sequence of the full-length follistatin-3 polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209199; and (f) the amino acid sequence of the mature follistatin-3 polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199. The polypeptides of the present invention also include polypeptides having an amino acid sequence at least 80% identical, more preferably at least 90% identical, and still more preferably 95%, 96%, 97%, 98% or 99% identical to those described in (a), (b), (c), (d), (e) or (f) above, as

well as polypeptides having an amino acid sequence with at least 90% similarity, and more preferably at least 95% similarity, to those above.

An additional embodiment of this aspect of the invention relates to a peptide or polypeptide which comprises the amino acid sequence of an epitope-bearing portion of a follistatin-3 polypeptide having an amino acid sequence described in (a), (b), (c), (d), (e) or (f) above. Peptides or polypeptides having the amino acid sequence of an epitope-bearing portion of a follistatin-3 polypeptide of the invention include portions of such polypeptides with at least six or seven, preferably at least nine, and more preferably at least about 30 amino acids to about 50 amino acids, although epitope-bearing polypeptides of any length up to and including the entire amino acid sequence of a polypeptide of the invention described above also are included in the invention.

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of a follistatin-3 polypeptide having an amino acid sequence which contains at least one amino acid substitution, but not more than 50 amino acid substitutions, even more preferably, not more than 40 amino acid substitutions, still more preferably, not more than 30 amino acid substitutions, and still even more preferably, not more than 20 amino acid substitutions. Of course, in order of ever-increasing preference, it is highly preferable for a peptide or polypeptide to have an amino acid sequence which comprises the amino acid sequence of a follistatin-3 polypeptide, which contains at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid substitutions. In specific embodiments, the number of additions, substitutions, and/or deletions in the amino acid sequence of Figures 1A and 1B, Figures 2A and 2B, or fragments thereof (e.g., the mature form and/or other fragments described herein), is 1-5, 5-10, 5-25, 5-50, 10-50 or 50-150, conservative amino acid substitutions are preferable.

In another embodiment, the invention provides an isolated antibody that binds specifically to a follistatin-3 polypeptide having an amino acid sequence described in (a), (b), (c), (d), (e) or (f) above. The invention further provides methods for isolating antibodies that bind specifically to a follistatin-3 polypeptide having an amino acid sequence as described herein. Such antibodies are useful diagnostically or therapeutically as described below.

The invention also provides for pharmaceutical compositions comprising follistatin-3 polypeptides, particularly human follistatin-3 polypeptides, which may be employed, for instance, to treat cancers and other cellular growth and differentiation disorders, as well as disorders of the reproductive system. Methods of treating individuals in need of follistatin-3 polypeptides are also provided.

The invention further provides compositions comprising a follistatin-3 polynucleotide or a follistatin-3 polypeptide for administration to cells *in vitro*, to cells *ex vivo* and to cells *in vivo*, or to a multicellular organism. In certain particularly preferred embodiments of this aspect of the invention, the compositions comprise a follistatin-3 polynucleotide for expression of a follistatin-3 polypeptide in a host organism for treatment of disease. Particularly preferred in this regard is expression in a human patient for treatment of a dysfunction associated with aberrant endogenous activity of follistatin-3.

The present invention also provides a screening method for identifying compounds capable of enhancing or inhibiting a biological activity of the follistatin-3 polypeptide, which involves contacting a ligand which is inhibited by the follistatin-3 polypeptide with the candidate compound in the presence of a follistatin-3 polypeptide, assaying receptor-binding activity of the ligand in the presence of the candidate compound and of follistatin-3 polypeptide, and comparing the ligand activity to a standard level of activity, the standard being assayed when contact is made between the ligand itself in the presence of the follistatin-3 polypeptide and the absence of the candidate compound. In this assay, an increase in ligand activity over the standard indicates that the candidate compound is an agonist of follistatin-3 activity and a decrease in ligand activity compared to the standard indicates that the compound is an antagonist of follistatin-3 activity.

In another aspect, a screening assay for agonists and antagonists is provided which involves determining the effect a candidate compound has on follistatin-3 binding to activin or an activin-like molecule. In particular, the method involves contacting the activin or an activin-like molecule with a follistatin-3 polypeptide and a candidate compound and determining whether follistatin-3 polypeptide binding to the activin or an activin-like molecule is increased or decreased due to the presence of the candidate

compound. In this assay, an increase in binding of follistatin-3 over the standard binding indicates that the candidate compound is an agonist of follistatin-3 binding activity and a decrease in follistatin-3 binding compared to the standard indicates that the compound is an antagonist of follistatin-3 binding activity.

It has been discovered that follistatin-3 is expressed not only in Hodgkin's Lymphoma but also in synovial fibroblasts, gall bladder, resting and serum-induced smooth muscle, testes, Merkel cells, HEL cells, hippocampus, TNF-alpha- and IFN-induced epithelial cells, keratinocyte, amygdala depression, HL-60 cells, hepatoma, progesterone-treated epidermal cells, endothelial cells, HSC172 cells, epithelioid sarcoma, activated T-cells, breast lymph node, pancreatic carcinoma, fetal dura mater, fetal lung, epididymis, placenta, dendritic cells, rejected kidney, and uterine cancer. Therefore, nucleic acids of the invention are useful as hybridization probes for differential identification of the tissue(s) or cell type(s) present in a biological sample. Similarly, polypeptides and antibodies directed to those polypeptides are useful to provide immunological probes for differential identification of the tissue(s) or cell type(s). In addition, for a number of disorders of the above tissues or cells, particularly of the reproductive system, or disorders of the regulation of cell growth and differentiation, significantly higher or lower levels of follistatin-3 gene expression may be detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) taken from an individual having such a disorder, relative to a "standard" follistatin-3 gene expression level, i.e., the follistatin-3 expression level in healthy tissue from an individual not having the reproductive system or regulation of cell growth and differentiation disorder. Thus, the invention provides a diagnostic method useful during diagnosis of such a disorder, which involves: (a) assaying follistatin-3 gene expression level in cells or body fluid of an individual; (b) comparing the follistatin-3 gene expression level with a standard follistatin-3 gene expression level, whereby an increase or decrease in the assayed follistatin-3 gene expression level compared to the standard expression level is indicative of disorder in the reproductive system or of a disorder of the regulation of cell growth and differentiation.

An additional aspect of the invention is related to a method for treating an individual in need of an increased level of follistatin-3 activity in the body comprising

administering to such an individual a composition comprising a therapeutically effective amount of an isolated follistatin-3 polypeptide of the invention or an agonist thereof.

A still further aspect of the invention is related to a method for treating an individual in need of a decreased level of follistatin-3 activity in the body comprising, administering to such an individual a composition comprising a therapeutically effective amount of an follistatin-3 antagonist. Preferred antagonists for use in the present invention are follistatin-3-specific antibodies.

Brief Description of the Figures

Figures 1A, 1B, and 1C show the nucleotide sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) of follistatin-3.

The predicted leader sequence of about 26 amino acids is underlined. Note that the methionine residue at the beginning of the leader sequence in Figure 1A is shown in position number (positive) 1, whereas the leader positions in the corresponding sequence of SEQ ID NO:2 are designated with negative position numbers. Thus, the leader sequence positions 1 to 26 in Figure 1A correspond to positions -26 to -1 in SEQ ID NO:2.

Two potential asparagine-linked glycosylation sites are marked in the amino acid sequence of follistatin-3. The sites are asparagine-73 and asparagine-215 in Figure 1A (asparagine-47 and asparagine-179 in SEQ ID NO:2), and are with the bold pound symbol (#) above the nucleotide sequence coupled with a bolded one letter abbreviation for the asparagine (N) in the amino acid sequence in Figure 1A; that is, the actual asparagine residues which are potentially glycosylated is bolded in Figure 1A. The potential N-linked glycosylation sequences are found at the following locations in the follistatin-3 amino acid sequence: N-73 through H-76 (N-73, L-74, T-75, H-76) and N-215 through Y-218 (N-215, V-216, T-217, Y-218). A potential Protein Kinase C (PKC) phosphorylation site is also marked in Figure 1A with a bolded tyrosine symbol (T) in the follistatin-3 amino acid sequence and an asterisk (*) above the first nucleotide encoding that tyrosine residue in the follistatin-3 nucleotide sequence. The potential PKC phosphorylation sequence is found in the follistatin-3 amino acid sequence from residue T-141 through residue R-143 (T-141, Y-142, R-143). Potential Casein Kinase II

(CK2) phosphorylation sites are also marked in Figure 1A with a bolded tyrosine or serine symbol (**T** or **S**) in the follistatin-3 amino acid sequence and an asterisk (*) above the first nucleotide encoding the appropriate tyrosine or serine residue in the follistatin-3 nucleotide sequence. Potential CK2 phosphorylation sequences are found at the following locations in the follistatin-3 amino acid sequence: T-57 through E-60 (T-57, R-58, A-59, E-60); T-141 through D-144 (T-141, Y-142, R-143, D-144); T-246 through E-249 (T-246, P-247, E-248, E-249); and S-255 through E-258 (S-255, A-256, E-257, E-258). Ten potential myristylation sites are found in the follistatin-3 amino acid sequence shown in Figure 1A. Potential myristylation sites are marked in Figure 1A with a double underline delineating the amino acid residues representing each potential myristolation site in the follistatin-3 amino acid sequence. The potential myristolation sites are located in the following postions in the follistatin-3 amino acid sequence: G-43 through C-48 (G-43, Q-44, E-45, A-46, T-47, C-48); G-65 through A-70 (G-65, N-66, I-67, D-68, T-69, A-70); G-78 through L-83 (G-78, N-79, K-80, I-81, N-82, L-83); G-88 through L-93 (G-88, L-89, V-90, H-91, C-92, L-93); G-136 through T-141 (G-136, S-137, D-138, G-139, A-140, T-141); G-188 through V-193 (G-188, S-189, A-190, H-191, C-192, V-193); G-207 through G-212 (G-207, Q-208, E-209, L-210, C-211, G-212); G-236 through G-241 (G-236, V-237, R-238, H-239, A-240, G-241); G-241 through T-246 (G-241, S-242, C-243, A-244, G-245, T-246); and G-252 through E-257 (G-252, G-253, E-254, S-255, A-256, E-257).

Figure 2 shows the regions of identity between the amino acid sequences of the follistatin-3 protein and translation product of the human mRNA for follistatin-1 (SEQ ID NO:3), determined by the computer program Bestfit (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) using the default parameters.

Figure 3 shows an analysis of the follistatin-3 amino acid sequence. Alpha, beta, turn and coil regions; hydrophilicity and hydrophobicity; amphipathic regions; flexible regions; antigenic index and surface probability, as predicted using default parameters, are shown.

In the "Antigenic Index or Jameson-Wolf" graph, the positive peaks indicate locations of the highly antigenic regions of the follistatin-3 protein, i.e., regions from

which epitope-bearing peptides of the invention can be obtained. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate follistatin-3-specific antibodies include: a polypeptide comprising amino acid residues from about Lys-54 to about Asp-62, from about Val-91 to about Leu-99, from about Lys-100 to about Gln-108, from about Cys-116 to about Pro-124, from about Gln-140 to about Leu-148, from about Trp-156 to about Ser-164, from about Arg-170, to about Gln-181, from about Cys-212 to about Phe-224, from about Tyr-239, to about Thr-247, from about Pro-251, to about Met-259, and from about Asp-263, to about His-271.

The data presented in Figure 3 are also represented in tabular form in Table I. The columns are labeled with the headings "Res", "Position", and Roman Numerals I-XIV. The column headings refer to the following features of the amino acid sequence presented in Figure 3 and Table I: "Res": amino acid residue of SEQ ID NO:2 or Figure 1A (which is the identical sequence shown in SEQ ID NO:2, with the exception that the residues are numbered 1-263 in Figure 1A and -18 through 348 in SEQ ID NO:4); "Position": position of the corresponding residue within SEQ ID NO:2 or Figures 2A and 2B (which is the identical sequence shown in SEQ ID NO:4, with the exception that the residues are numbered 1-366 in Figures 2A and 2B and -18 through 348 in SEQ ID NO:4); I: Alpha, Regions - Garnier-Robson; II: Alpha, Regions - Chou-Fasman; III: Beta, Regions - Garnier-Robson; IV: Beta, Regions - Chou-Fasman; V: Turn, Regions - Garnier-Robson; VI: Turn, Regions - Chou-Fasman; VII: Coil, Regions - Garnier-Robson; VIII: Hydrophilicity Plot - Kyte-Doolittle; IX: Hydrophobicity Plot - Hopp-Woods; X: Alpha, Amphipathic Regions - Eisenberg; XI: Beta, Amphipathic Regions - Eisenberg; XII: Flexible Regions - Karplus-Schulz; XIII: Antigenic Index - Jameson-Wolf; and XIV: Surface Probability Plot - Emini.

Detailed Description

The present invention provides isolated nucleic acid molecules comprising a polynucleotide encoding a follistatin-3 polypeptide having the amino acid sequence shown in SEQ ID NO:2, which was determined by sequencing a cloned cDNA. The nucleotide sequence shown in Figures 1A, 1B, and 1C (SEQ ID NO:1) was obtained by sequencing the HDTAH85 clone, which was deposited on August 8, 1997 at the

American Type Culture Collection, 10801 University Boulevard, Manassas, Virginia 20110-2209, and given accession number ATCC 209199. The deposited clone is contained in the pBluescript SK(-) plasmid (Stratagene, La Jolla, CA).

The follistatin-3 protein of the present invention shares sequence homology with the translation product of the human mRNA for follistatin-1 (Figure 2; SEQ ID NO:3). Follistatin-1 is thought to be an important factor in the regulation of follicle development and spermatogenesis in the reproductive systems. Follistatin-1 acts as an antagonist of activin by stoichiometrically binding to activin and preventing interaction with the activin receptor. It is thought that, in addition to activin, follistatin-1 may act in a similar manner by targeting additional members of the TGF-beta superfamily.

Nucleic Acid Molecules

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the Model 373 from Applied Biosystems, Inc., Foster City, CA), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

By "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides,

and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U), where each thymidine deoxyribonucleotide (T) in the specified deoxyribonucleotide sequence is replaced by the ribonucleotide uridine (U).

Using the information provided herein, such as the nucleotide sequence in Figures 1A, 1B, and 1C (SEQ ID NO:1), a nucleic acid molecule of the present invention encoding a follistatin-3 polypeptide may be obtained using standard cloning and screening procedures, such as those for cloning cDNAs using mRNA as starting material. Illustrative of the invention, the nucleic acid molecule described in Figures 1A, 1B, and 1C (SEQ ID NO:1) was discovered in a cDNA library derived from Hodgkin's Lymphoma.

Additional clones of the same gene were also identified in cDNA libraries from the following cells and tissues: synovial fibroblasts, gall bladder, resting and serum-induced smooth muscle, testes, Merkel cells, HEL cells, hippocampus, TNF-alpha- and IFN-induced epithelial cells, keratinocyte, amygdala depression, HL-60 cells, hepatoma, progesterone-treated epidermal cells, endothelial cells, HSC172 cells, epithelioid sarcoma, activated T-cells, breast lymph node, pancreatic carcinoma, fetal dura mater, fetal lung, epididymis, placenta, dendritic cells, rejected kidney, and uterine cancer.

The determined nucleotide sequence of the follistatin-3 cDNA of Figures 1A, 1B, and 1C (SEQ ID NO:1) contains an open reading frame encoding a protein of 263 amino acid residues, with an initiation codon at nucleotide positions 19-21 of the nucleotide sequence in Figure 1A (SEQ ID NO:1), and a deduced molecular weight of about 27.7 kDa. The amino acid sequence of the follistatin-3 protein shown in SEQ ID NO:2 is about 43.2% identical to human mRNA for follistatin-1 (Figure 2, Shimasaki, S., *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**:4218-4222 (1988); GenBank Accession No. J03771).

The open reading frame of the follistatin-3 gene shares sequence homology with the translation product of the human mRNA for follistatin-1 (Figure 2, SEQ ID NO:3). The homology between follistatin-1 and follistatin-3 indicates that follistatin-3 may also be involved in a physiological regulation of cell growth and differentiation, particularly with regard to cells of the reproductive system.

As one of ordinary skill would appreciate, due to the possibilities of sequencing errors discussed above, the actual complete follistatin-3 polypeptide encoded by the deposited cDNA, which comprises about 263 amino acids, may be somewhat longer or shorter. More generally, the actual open reading frame may be anywhere in the range of ±20 amino acids, more likely in the range of ±10 amino acids, of that predicted from either the methionine codon from the N-terminus shown in Figure 1A (SEQ ID NO:1). It will further be appreciated that, depending on the analytical criteria used for identifying various functional domains, the exact "address" of the mature form of the follistatin-3 polypeptide may differ slightly from the predicted positions above. For example, the exact location of the cleavage site of the precursor form of the mature follistatin-3 molecule shown in SEQ ID NO:2 may vary slightly (e.g., the address may "shift" by about 6 residues, depending on the criteria used to define the cleavage site. In this case, the ends of the signal peptide and the beginning of the mature follistatin-3 molecule were predicted using the HGSI SignalP computer algorithm. One of skill in the art will realize that another widely accepted computer algorithm used to predict potential sites of polypeptide cleavage, PSORT, will predict the cleavage of an N-terminal signal peptide from the follistatin-3 polypeptide at a point slightly different from that predicted by the HGSI SignalP algorithm. In either case, as discussed further below, the invention further provides polypeptides having various residues deleted from the N-terminus of the complete polypeptide, including polypeptides corresponding to either of the predicted mature follistatin-3 polypeptides described herein.

The amino acid sequence of the complete follistatin-3 protein includes a leader sequence and a mature protein, as shown in SEQ ID NO:2. More in particular, the present invention provides nucleic acid molecules encoding a mature form of the follistatin-3 protein. Thus, according to the signal hypothesis, once export of the growing protein chain across the rough endoplasmic reticulum has been initiated, proteins secreted by mammalian cells have a signal or secretory leader sequence which is cleaved from the complete polypeptide to produce a secreted "mature" form of the protein. Most mammalian cells and even insect cells cleave secreted proteins with the same specificity. However, in some cases, cleavage of a secreted protein is not entirely uniform, which results in two or more mature species of the protein. Further, it has long

been known that the cleavage specificity of a secreted protein is ultimately determined by the primary structure of the complete protein, that is, it is inherent in the amino acid sequence of the polypeptide. Therefore, the present invention provides a nucleotide sequence encoding the mature follistatin-3 polypeptide having the amino acid sequence encoded by the cDNA clone contained in the host identified as ATCC Deposit No. 209199. By the "mature follistatin-3 polypeptide having the amino acid sequence encoded by the cDNA clone in ATCC Deposit No. 209199" is meant the mature form(s) of the follistatin-3 protein produced by expression in a mammalian cell (e.g., COS cells, as described below) of the complete open reading frame encoded by the human DNA sequence of the clone contained in the vector in the deposited host.

In addition, methods for predicting whether a protein has a secretory leader as well as the cleavage point for that leader sequence are available. For instance, the method of McGeoch (*Virus Res.* **3**:271-286 (1985)) uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje (*Nucleic Acids Res.* **14**:4683-4690 (1986)) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2 where +1 indicates the amino terminus of the mature protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80% (von Heinje, *supra*). However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the complete follistatin-3 polypeptide was analyzed by a variation of the computer program "PSORT", available from Dr. Kenta Nakai of the Institute for Chemical Research, Kyoto University (Nakai, K. and Kanehisa, M. *Genomics* **14**:897-911 (1992)), which is an expert system for predicting the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. Thus, the computation analysis above predicted a single cleavage site within the complete amino acid sequence shown in SEQ ID NO:2 (see above discussion).

As indicated, nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and

genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. For example, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Isolated nucleic acid molecules of the present invention include DNA molecules comprising an open reading frame (ORF) with an initiation codon at positions 19-21 of the nucleotide sequence shown in Figure 1A (SEQ ID NO:1).

Also included are DNA molecules comprising the coding sequence for the predicted mature follistatin-3 protein shown at positions 1-237 of SEQ ID NO:2.

In addition, isolated nucleic acid molecules of the invention include DNA molecules which comprise a sequence substantially different from those described above but which, due to the degeneracy of the genetic code, still encode the follistatin-3 protein. Of course, the genetic code and species-specific codon preferences are well known in the art. Thus, it would be routine for one skilled in the art to generate the degenerate variants described above, for instance, to optimize codon expression for a particular host (e.g., change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

In another aspect, the invention provides isolated nucleic acid molecules encoding the follistatin-3 polypeptide having an amino acid sequence encoded by the cDNA clone contained in the plasmid deposited as ATCC Deposit No. 209199 on August 8, 1997.

Preferably, this nucleic acid molecule will encode the mature polypeptide encoded by the above-described deposited cDNA clone.

The invention further provides an isolated nucleic acid molecule having the nucleotide sequence shown in Figures 1A, 1B, and 1C (SEQ ID NO:1) or the nucleotide sequence of the follistatin-3 cDNA contained in the above-described deposited clone, or a nucleic acid molecule having a sequence complementary to one of the above sequences. Such isolated molecules, particularly DNA molecules, are useful as probes for gene mapping, by *in situ* hybridization with chromosomes, and for detecting expression of the follistatin-3 gene in human tissue, for instance, by Northern blot analysis.

The present invention is further directed to nucleic acid molecules encoding portions of the nucleotide sequences described herein as well as to fragments of the isolated nucleic acid molecules described herein. In particular, the invention provides a polynucleotide having a nucleotide sequence representing the portion of SEQ ID NO:1 which consists of positions 1-810 of SEQ ID NO:1.

In addition, the invention provides nucleic acid molecules having nucleotide sequences related to extensive portions of SEQ ID NO:1 which have been determined from the following related cDNA clones: HHPDX66R (SEQ ID NO:4), HDTAH61R (SEQ ID NO:5), HSBAV55R (SEQ ID NO:6), HUKFS32R (SEQ ID NO:7), HOOAD78R (SEQ ID NO:8), HAQAG52R (SEQ ID NO:9), HTLEJ56R (SEQ ID NO:10), HLMNX90R (SEQ ID NO:11).

Further, the invention includes a polynucleotide comprising any portion of at least about 30 nucleotides, preferably at least about 50 nucleotides, of SEQ ID NO:1 from residue 1 to 500. More preferably, the invention includes a polynucleotide comprising nucleotide residues 100-500, 200-500, 300-500, 400-500, 100-400, 200-400, 300-400, 100-300, 200-300, 100-200, 100-2495, 250-2495, 500-2495, 1000-2495, 1500-2495, 2000-2495, 100-2000, 250-2000, 500-2000, 1000-2000, 1500-2000, 100-1500, 250-1500, 500-1500, 1000-1500, 100-1000, 250-1000, and 500-1000.

More generally, by a fragment of an isolated nucleic acid molecule having the nucleotide sequence of the deposited cDNA or the nucleotide sequence shown in Figures 1A, 1B, and 1C (SEQ ID NO:1) is intended fragments at least about 15 nt, and more

preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length which are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments 50-300 nt in length are also useful according to the present invention as are fragments corresponding to most, if not all, of the nucleotide sequence of the deposited cDNA or as shown in Figures 1A, 1B, and 1C (SEQ ID NO:1). By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or more contiguous bases from the nucleotide sequence of the deposited cDNA or the nucleotide sequence as shown in Figures 1A, 1B, and 1C (SEQ ID NO:1). Preferred nucleic acid fragments of the present invention include nucleic acid molecules encoding epitope-bearing portions of the follistatin-3 polypeptide as identified in Figure 3 and described in more detail below.

In specific embodiments, the polynucleotide fragments of the invention encode a polypeptide which demonstrates a functional activity. By a polypeptide demonstrating "functional activity" is meant, a polypeptide capable of displaying one or more known functional activities associated with a complete, mature or active form of the follistatin-3 polypeptide. Such functional activities include, but are not limited to, biological activity ((e.g., modulating the follicle stimulating hormone (FSH) synthetic pathway, increasing estradiol production, binding activin, stimulator of gonadotropin biosynthesis and secretion, regulator of ovarian and placental steroidogenesis, and oocyte and spermatogonial maturation factor)), antigenicity [ability to bind (or compete with a follistatin-3 polypeptide for binding) to an anti-follistatin-3 antibody], immunogenicity (ability to generate antibody which binds to a follistatin-3 polypeptide), the ability to form polymers with other follistatin-3 or inhibin or TGF-beta polypeptides, and ability to bind to a receptor or ligand for a follistatin-3 polypeptide.

Preferred nucleic acid fragments of the present invention also include nucleic acid molecules encoding one or more of the following domains of follistatin-3: amino acid residues 7-16, 34-45, 78-86, 91-100, 108-122, 131-145, 156-169, 184-192, and 196-210 of SEQ ID NO:2.

In specific embodiments, the polynucleotide fragments of the invention encode antigenic regions. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate follistatin-3-specific antibodies include: a polypeptide comprising

amino acid residues from about Leu-14 to about Ala-20, from about Ser-46 to about Ile-55, from about Gly-88 to about Pro-97, from about Gly-113 to about Leu-133, from about Arg-138 to about Glu-146, from about Pro-177 to about Thr-191, from about Gly-219 to about Val-237.

In additional embodiments, the polynucleotides of the invention encode functional attributes of follistatin-3. Preferred embodiments of the invention in this regard include fragments that comprise alpha-helix and alpha-helix forming regions ("alpha-regions"), beta-sheet and beta-sheet forming regions ("beta-regions"), turn and turn-forming regions ("turn-regions"), coil and coil-forming regions ("coil-regions"), hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions and high antigenic index regions of follistatin-3.

The data representing the structural or functional attributes of follistatin-3 set forth in Figure 3 and/or Table I, as described above, was generated using the various modules and algorithms of the DNA*STAR set on default parameters. In a preferred embodiment, the data presented in columns VIII, IX, XIII, and XIV of Table I can be used to determine regions of follistatin-3 which exhibit a high degree of potential for antigenicity. Regions of high antigenicity are determined from the data presented in columns VIII, IX, XIII, and/or IV by choosing values which represent regions of the polypeptide which are likely to be exposed on the surface of the polypeptide in an environment in which antigen recognition may occur in the process of initiation of an immune response.

Certain preferred regions in these regards are set out in Figure 3, but may, as shown in Table I, be represented or identified by using tabular representations of the data presented in Figure 3. The DNA*STAR computer algorithm used to generate Figure 3 (set on the original default parameters) was used to present the data in Figure 3 in a tabular format (*See* Table I). The tabular format of the data in Figure 3 may be used to easily determine specific boundaries of a preferred region.

The above-mentioned preferred regions set out in Figure 3 and in Table I include, but are not limited to, regions of the aforementioned types identified by analysis of the amino acid sequence set out in Figures 1A, 1B, and 1C. As set out in Figure 3 and in

Table I, such preferred regions include Garnier-Robson alpha-regions, beta-regions, turn-regions, and coil-regions, Chou-Fasman alpha-regions, beta-regions, and coil-regions, Kyte-Doolittle hydrophilic regions and hydrophobic regions, Eisenberg alpha- and beta-amphipathic regions, Karplus-Schulz flexible regions, Emini surface-forming regions and Jameson-Wolf regions of high antigenic index.

Table I

	Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
	Met	1	.	.	B	.	.	.	.	0.31	-0.24	*	*	.	1.07	1.11
	Arg	2	.	.	B	.	.	.	.	0.49	-0.17	*	*	.	1.13	0.88
	Pro	3	.	.	.	.	T	.	.	0.53	-0.17	*	*	.	1.89	1.06
	Gly	4	.	.	.	.	T	.	.	0.71	-0.17	*	*	.	2.10	1.06
	Ala	5	.	.	.	.	.	T	C	0.29	-0.36	.	*	F	1.89	0.84
	Pro	6	.	.	.	.	.	T	C	0.60	0.33	.	*	F	1.08	0.45
	Gly	7	.	.	.	.	.	T	C	0.28	0.81	.	*	F	0.57	0.48
	Pro	8	.	.	B	.	.	T	.	-0.32	0.81	.	.	F	0.16	0.73
	Leu	9	.	.	B	.	.	.	.	-0.19	1.80	.	.	F	-0.25	0.39
	Trp	10	.	.	B	.	.	.	.	0.11	1.06	.	.	.	-0.40	0.61
	Pro	11	.	.	B	.	.	.	.	-0.02	1.49	.	.	.	-0.40	0.41
	Leu	12	.	.	B	.	.	T	.	-0.27	1.49	.	.	.	-0.20	0.49
	Pro	13	.	.	.	.	T	T	.	-0.87	1.30	.	.	.	0.20	0.48
	Trp	14	.	.	.	.	T	T	.	-0.64	1.07	.	.	.	0.20	0.25
	Gly	15	.	.	.	.	.	T	C	-0.64	1.14	.	.	.	0.00	0.31
	Ala	16	.	A	.	.	.	.	C	-1.02	1.37	.	.	.	-0.40	0.21
	Leu	17	.	A	B	.	.	.	.	-1.07	1.44	.	.	.	-0.60	0.20
	Ala	18	.	A	B	B	.	.	.	-1.20	1.17	.	.	.	-0.60	0.15
	Trp	19	.	A	B	B	.	.	.	-1.61	1.17	.	.	.	-0.60	0.15
	Ala	20	.	A	B	B	.	.	.	-2.12	1.46	.	.	.	-0.69	0.16
	Val	21	.	A	B	B	.	.	.	-1.83	1.43	.	.	.	-0.60	0.11
	Gly	22	.	A	B	B	.	.	.	-1.32	1.30	.	.	.	-0.60	0.15
	Phe	23	.	.	B	B	.	.	.	-1.33	0.77	.	.	.	-0.60	0.19
	Val	24	.	.	B	B	.	.	.	-1.39	0.89	.	.	.	-0.60	0.26
	Ser	25	.	.	B	.	.	.	.	-1.10	0.67	*	.	.	-0.40	0.26
	Ser	26	.	.	B	.	.	.	.	-0.59	0.63	.	.	F	-0.25	0.40
	Met	27	.	.	.	.	T	.	.	-0.24	0.27	.	.	F	0.45	0.53
	Gly	28	.	.	.	.	T	T	.	0.24	0.03	.	.	F	0.82	0.64
	Ser	29	.	.	.	.	T	T	.	0.51	0.07	.	.	F	0.99	0.74
	Gly	30	.	.	.	.	.	T	C	0.60	0.19	.	.	F	0.96	0.76
	Asn	31	.	.	.	.	.	T	C	0.36	-0.00	.	.	F	1.88	1.18
	Pro	32	.	.	.	.	.	.	C	0.81	-0.00	.	.	F	1.70	0.87
	Asn	33	.	.	.	.	.	T	C	0.30	0.04	.	.	F	1.13	0.87
	Pro	34	.	.	.	.	T	T	.	-0.07	0.26	.	.	F	1.16	0.40
	Gly	35	.	.	.	.	T	T	.	-0.01	0.43	*	.	F	0.69	0.14
	Gly	36	.	.	B	.	.	T	.	-0.82	0.91	*	.	F	0.12	0.15
	Val	37	.	A	B	.	.	.	.	-0.61	1.10	*	.	.	-0.60	0.68
	Cys	38	.	A	B	.	.	.	.	-0.02	1.07	*	.	.	-0.60	0.14
	Trp	39	.	A	B	.	.	.	.	-0.16	1.08	*	.	.	-0.60	0.24
	Leu	40	.	A	B	.	.	.	.	0.19	1.04	*	.	.	-0.32	0.32
	Gln	41	.	.	B	.	.	T	.	0.53	0.80	*	.	F	0.66	1.02
	Gln	42	.	.	.	.	T	T	.	0.80	0.23	*	.	F	1.64	1.68
	Gly	43	.	.	.	.	T	T	.	1.16	-0.19	*	.	F	2.52	2.06
	Gln	44	.	.	.	.	T	T	.	0.78	-0.39	*	.	F	2.80	1.72
	Glu	45	.	.	.	.	T	.	.	1.29	-0.21	*	.	F	2.17	0.53
	Ala	46	.	.	.	.	T	T	.	0.48	-0.23	*	.	F	2.09	0.72
	Thr	47	.	.	B	.	.	T	.	-0.38	0.03	.	.	.	0.66	0.34
	Cys	48	.	.	B	.	.	T	.	-0.84	0.27	.	.	.	0.38	0.15
	Ser	49	.	.	B	.	.	T	.	-0.84	0.96	.	.	.	-0.20	0.12
	Leu	50	.	.	B	B	.	.	.	-1.16	0.86	.	*	.	-0.60	0.14
55	Val	51	.	.	B	B	.	.	.	-0.57	0.86	.	*	.	-0.60	0.39
	Leu	52	.	.	B	B	.	.	.	-1.11	0.29	.	*	.	-0.30	0.48
	Gln	53	.	.	B	B	.	.	.	-0.76	0.54	*	*	F	-0.43	0.43
	Thr	54	.	.	B	B	.	.	.	-0.34	0.34	*	.	F	-0.15	0.85
	Asp	55	.	.	B	B	.	.	.	-0.12	-0.30	*	*	F	0.60	2.01
60	Val	56	.	A	B	B	.	.	.	0.73	-0.49	*	.	F	0.60	1.17

Table 1 (continued)

	Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
	Thr	57	.	A	B	B	.	.	.	0.88	-0.89	*	*	F	0.90	1.41
	Arg	58	.	A	B	B	.	.	.	0.21	-0.80	*	*	F	0.75	0.45
	Ala	59	.	A	B	B	.	.	.	-0.07	-0.23	*	*	.	0.30	0.33
	Glu	60	.	A	B	B	.	.	.	-0.37	-0.37	*	*	.	0.30	0.23
	Cys	61	.	A	B	.	.	.	.	0.14	-0.47	*	*	.	0.55	0.16
	Cys	62	.	.	.	.	T	T	.	0.46	-0.04	*	*	.	1.60	0.15
	Ala	63	.	.	.	.	T	T	.	-0.54	-0.14	*	*	.	1.85	0.14
	Ser	64	.	.	.	.	T	T	.	0.04	0.54	.	*	F	1.35	0.19
	Gly	65	.	.	.	.	T	T	.	-0.27	-0.03	.	*	F	2.50	0.58
	Asn	66	.	.	.	.	T	T	.	-0.19	-0.11	.	*	F	2.25	0.83
	Ile	67	.	.	B	.	.	T	.	0.19	-0.11	*	*	F	1.60	0.62
	Asp	68	.	.	B	.	.	T	.	0.48	0.41	*	*	F	0.45	0.66
	Thr	69	.	.	B	.	.	T	.	0.78	0.37	*	*	F	0.50	0.55
	Ala	70	.	.	B	.	.	.	.	0.31	0.37	*	*	.	0.05	1.26
	Trp	71	.	.	B	.	.	T	.	-0.00	0.37	*	.	.	0.10	0.62
	Ser	72	.	.	B	.	.	T	.	0.86	0.86	*	.	.	-0.20	0.62
	Asn	73	.	.	B	.	.	T	.	0.64	0.87	*	.	.	-0.20	0.84
	Leu	74	.	.	.	.	.	T	C	0.61	0.80	*	.	.	0.43	1.24
	Thr	75	.	.	.	.	.	.	C	1.20	0.31	*	.	.	0.66	0.91
	His	76	.	.	.	.	.	T	C	1.53	0.33	*	.	F	1.29	0.91
	Pro	77	.	.	.	.	.	T	C	0.94	-0.07	*	.	F	2.32	2.22
	Gly	78	.	.	.	.	T	T	.	0.94	-0.07	*	*	F	2.80	1.08
	Asn	79	.	.	.	.	T	T	.	0.94	-0.16	*	.	F	2.52	1.27
	Lys	80	.	.	B	.	.	.	.	0.46	0.03	*	*	F	0.89	0.68
	Ile	81	.	.	B	.	.	.	.	0.13	0.29	.	.	F	0.61	0.57
	Asn	82	.	.	B	.	.	.	.	-0.36	0.29	.	*	.	0.18	0.35
	Leu	83	.	.	B	B	.	.	.	-0.82	0.67	.	.	.	-0.60	0.15
	Leu	84	.	.	B	B	.	.	.	-1.17	1.36	.	*	.	-0.60	0.18
	Gly	85	.	.	B	B	.	.	.	-2.02	1.10	.	*	.	-0.60	0.11
	Phe	86	.	.	B	B	.	.	.	-1.99	1.39	.	.	.	-0.60	0.11
	Leu	87	.	.	B	B	.	.	.	-2.02	1.34	.	.	.	-0.60	0.10
	Gly	88	.	.	B	B	.	.	.	-1.88	1.16	*	.	.	-0.60	0.13
	Leu	89	.	.	B	B	.	.	.	-1.88	1.30	.	.	.	-0.60	0.08
	Val	90	.	.	B	B	.	.	.	-1.74	1.20	*	.	.	-0.60	0.08
	His	91	.	.	B	B	.	.	.	-1.71	0.94	*	.	.	-0.60	0.13
	Cys	92	.	.	B	B	.	.	.	-0.86	1.09	.	.	.	-0.60	0.08
	Leu	93	.	.	B	B	.	.	.	-0.51	0.40	.	.	.	0.01	0.23
	Pro	94	.	.	.	B	T	.	.	-0.00	-0.24	.	.	.	1.32	0.28
	Cys	95	.	.	.	.	T	T	.	0.19	-0.36	.	.	.	2.03	0.70
	Lys	96	.	.	.	.	T	T	.	0.22	-0.36	.	.	F	2.49	0.46
	Asp	97	.	.	.	.	T	T	.	0.54	-1.04	*	.	F	3.10	0.49
	Ser	98	.	.	.	.	T	T	.	0.50	-1.04	*	.	F	2.79	0.91
	Cys	99	.	.	.	.	T	T	.	0.71	-0.97	*	.	F	2.48	0.34
	Asp	100	.	.	B	.	.	T	.	0.71	-0.97	*	.	F	1.77	0.35
	Gly	101	.	.	B	.	.	T	.	0.32	-0.40	*	.	F	1.47	0.14
	Val	102	.	.	B	.	.	T	.	0.11	-0.36	*	.	.	1.32	0.26
	Glu	103	.	.	B	.	.	.	.	0.07	-0.50	*	.	.	1.73	0.24
	Cys	104	.	.	.	.	T	.	.	0.78	-0.07	*	.	F	2.29	0.24
	Gly	105	.	.	.	.	T	T	.	0.19	-0.50	*	.	F	3.10	0.64
	Pro	106	.	.	.	.	T	T	.	-0.13	-0.64	*	.	F	2.79	0.38
	Gly	107	.	.	.	.	T	T	.	0.83	-0.07	*	.	F	2.38	0.38
	Lys	108	.	.	.	.	T	T	.	0.23	-0.64	*	.	F	2.17	0.74
	Ala	109	.	A	B	.	.	.	.	0.09	-0.46	*	.	.	0.61	0.48
	Cys	110	.	A	B	.	.	.	.	0.09	-0.20	*	.	.	0.30	0.40
	Arg	111	.	A	B	.	.	.	.	-0.04	-0.20	*	.	.	0.30	0.20
	Met	112	.	A	B	.	.	.	.	0.41	0.23	*	.	.	-0.30	0.19

Table I (continued)

	Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
5	Leu	113	.	A	.	.	T	.	.	0.16	-0.27	*	*	.	1.04	0.70
	Gly	114	.	A	.	.	T	.	.	0.86	-0.41	*	*	F	1.53	0.55
	Gly	115	.	.	.	.	T	.	.	0.86	-0.41	*	*	F	2.22	1.10
	Arg	116	.	.	.	.	.	T	C	0.74	-0.46	*	*	F	2.41	0.71
	Pro	117	.	.	.	.	T	T	.	0.68	-1.14	.	*	F	3.40	1.25
10	Arg	118	.	.	.	.	T	T	.	0.90	-1.00	.	*	F	2.91	0.68
	Cys	119	.	.	B	.	.	T	.	1.03	-0.93	.	*	.	2.02	0.35
	Glu	120	.	.	B	.	.	.	.	1.38	-0.50	.	*	.	1.73	0.35
	Cys	121	.	.	B	.	.	.	.	0.60	-0.93	.	*	.	1.64	0.30
	Ala	122	.	.	B	.	.	T	.	0.51	-0.36	.	*	.	1.45	0.30
15	Pro	123	.	.	.	.	T	T	.	0.06	-0.54	.	*	P	2.35	0.23
	Asp	124	.	.	.	.	T	T	.	-0.09	-0.13	.	.	F	2.59	0.43
	Cys	125	.	.	.	.	T	T	.	-0.30	-0.60	.	.	F	2.25	0.33
	Ser	126	.	.	.	.	T	.	.	-0.22	-0.87	*	*	F	1.89	0.35
	Gly	127	.	.	.	.	T	.	.	0.48	-0.80	*	*	F	1.55	0.21
20	Leu	128	.	.	B	.	.	.	.	-0.12	-0.60	*	*	.	0.75	0.77
	Pro	129	.	.	B	.	.	.	.	-0.12	0.11	.	*	.	-0.10	0.47
	Ala	130	.	.	B	.	.	.	.	-0.31	0.13	.	*	.	-0.19	0.83
	Arg	131	.	.	B	B	.	.	.	-0.68	0.34	.	*	.	-0.30	0.74
	Leu	132	.	.	B	B	.	.	.	-0.68	0.23	.	*	.	-0.39	0.26
25	Gln	133	.	.	B	B	.	.	.	-0.17	0.23	*	*	.	-0.39	0.25
	Val	134	.	.	B	B	.	.	.	0.04	0.11	*	*	.	-0.30	0.17
	Cys	135	.	.	B	B	.	.	.	0.29	0.11	*	*	.	-0.02	0.35
	Gly	136	.	.	B	.	.	T	.	-0.41	-0.14	*	*	F	1.41	0.20
	Ser	137	.	.	.	.	T	T	.	0.09	-0.04	.	.	F	2.09	0.27
30	Asp	138	.	.	.	.	T	T	.	-0.16	-0.20	.	*	F	2.37	0.73
	Gly	139	.	.	.	.	T	T	.	0.81	-0.01	.	.	F	2.80	1.16
	Ala	140	.	.	.	.	T	.	.	1.48	-0.46	.	.	F	2.32	1.70
	Thr	141	.	.	B	.	.	.	.	1.82	-0.83	.	.	.	1.99	1.70
	Tyr	142	.	.	B	.	.	T	.	1.46	-0.83	.	*	.	2.11	2.97
35	Arg	143	.	.	B	.	.	T	.	1.46	-0.69	.	*	F	2.18	1.57
	Asp	144	.	.	B	.	.	T	.	0.99	-1.19	.	*	F	2.16	1.89
	Glu	145	.	.	B	.	.	T	.	1.69	-0.99	.	*	.	2.60	0.99
	Cys	146	.	A	B	.	.	.	.	1.41	-1.74	.	*	.	1.40	0.99
	Glu	147	A	A	.	.	.	.	.	1.07	-1.24	.	*	.	1.20	0.60
40	Leu	148	A	A	.	.	.	.	.	1.07	-0.74	.	*	.	1.00	0.35
	Arg	149	A	A	.	.	.	.	.	0.40	-0.74	.	*	.	0.95	1.28
	Ala	150	A	A	.	.	.	.	.	0.51	-0.74	*	*	.	0.60	0.40
	Ala	151	.	A	.	.	T	.	.	0.83	-0.74	.	*	.	1.00	0.94
	Arg	152	.	A	.	.	T	.	.	0.80	-1.08	.	*	.	1.00	0.48
45	Cys	153	.	A	.	.	T	.	.	1.40	-0.58	.	*	.	1.27	0.64
	Arg	154	.	A	.	.	T	.	.	1.29	-0.57	.	*	.	1.54	0.98
	Gly	155	.	A	.	.	T	.	.	1.07	-1.07	.	*	F	1.96	0.84
	His	156	.	.	.	.	.	T	C	1.36	-0.39	.	*	F	2.28	1.29
	Pro	157	.	.	.	.	.	T	C	0.39	-0.57	.	*	F	2.70	0.88
50	Asp	158	.	.	.	.	T	T	.	0.46	0.07	*	*	F	1.73	0.66
	Leu	159	.	.	B	.	.	T	.	0.19	0.26	*	*	.	0.91	0.48
	Ser	160	.	.	B	B	.	.	.	0.56	0.51	*	*	.	-0.86	0.49
	Val	161	.	.	B	B	.	.	.	0.24	0.09	.	*	.	-0.03	0.57
	Met	162	.	.	B	B	.	.	.	0.57	0.51	.	*	.	-0.26	0.69
55	Tyr	163	.	.	B	.	.	T	.	-0.18	-0.17	*	*	.	1.53	1.00
	Arg	164	.	.	B	.	.	T	.	0.82	0.01	.	*	.	1.12	0.72
	Gly	165	.	.	.	.	T	T	.	1.17	-0.63	.	*	F	3.06	1.43
	Arg	166	.	.	.	.	T	T	.	1.72	-1.24	.	*	F	3.42	1.83
	Cys	167	.	.	.	.	T	.	.	1.66	-1.61	*	*	F	2.86	1.35
60	Arg	168	.	.	.	.	T	T	.	1.90	-1.04	*	*	F	2.57	0.68

Table I (continued)

	Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
	Lys	169	.	.	.	.	T	T	.	1.76	-1.47	*	*	F	2.23	0.60
	Ser	170	.	.	.	.	T	T	.	1.28	-0.97	*	*	F	2.04	1.52
	Cys	171	.	.	.	.	T	T	.	0.28	-0.90	*	*	.	1.40	0.58
	Glu	172	.	.	B	B	.	.	.	0.28	-0.26	*	.	.	0.30	0.21
	His	173	.	.	B	B	.	.	.	-0.04	0.31	.	.	.	-0.30	0.09
	Val	174	.	.	B	B	.	.	.	0.02	0.36	.	*	.	-0.02	0.25
	Val	175	.	.	B	B	.	.	.	0.11	-0.21	.	*	.	0.86	0.38
	Cys	176	.	.	B	.	.	T	.	0.78	0.21	.	*	.	0.94	0.32
	Pro	177	.	.	.	.	T	T	.	0.48	0.11	.	*	F	1.77	0.74
	Arg	178	.	.	.	.	T	T	.	-0.16	-0.14	.	*	F	2.89	1.34
	Pro	179	.	.	.	.	T	T	.	-0.16	-0.21	.	*	F	2.52	1.34
	Gln	180	.	.	.	B	T	.	.	-0.16	-0.14	*	*	F	1.69	0.66
	Ser	181	.	.	B	B	.	.	.	0.51	0.07	*	*	F	0.41	0.24
	Cys	182	.	.	B	B	.	.	.	0.72	0.07	*	*	.	-0.02	0.26
	Val	183	.	.	B	B	.	.	.	0.30	0.04	*	*	.	-0.30	0.26
	Val	184	.	.	B	B	.	.	.	0.17	0.13	.	.	.	-0.02	0.28
	Asp	185	.	.	B	B	.	.	.	-0.13	0.17	.	.	F	0.41	0.52
	Gln	186	.	.	B	.	.	T	.	-0.42	-0.81	.	.	F	1.69	0.94
	Thr	187	.	.	.	.	T	T	.	0.21	-0.16	.	.	F	2.52	1.28
	Gly	188	.	.	.	.	T	T	.	0.40	-0.30	.	.	F	2.80	1.04
	Ser	189	.	.	.	.	T	T	.	0.40	0.27	.	.	F	1.77	0.32
	Ala	190	.	.	B	B	T	.	.	-0.46	0.51	.	.	.	0.64	0.17
	His	191	.	.	B	B	.	.	.	-1.12	0.67	*	.	.	-0.04	0.12
	Cys	192	.	.	B	B	.	.	.	-0.70	0.81	*	*	.	-0.32	0.05
	Val	193	.	.	B	B	.	.	.	-0.94	0.43	*	.	.	-0.60	0.10
	Val	194	.	.	B	B	.	.	.	-1.23	0.43	*	.	.	-0.60	0.07
	Cys	195	.	.	B	B	.	.	.	-0.86	0.43	*	.	.	-0.60	0.14
	Arg	196	.	.	B	B	.	.	.	-1.49	0.29	*	.	.	-0.30	0.28
	Ala	197	.	.	B	B	.	.	.	-1.03	0.21	*	.	.	-0.30	0.39
	Ala	198	.	.	B	.	.	T	.	-1.03	-0.00	*	.	.	0.79	0.59
	Pro	199	.	.	B	.	.	T	.	-0.39	0.07	*	.	.	0.10	0.22
	Cys	200	.	.	B	.	.	T	.	-0.02	0.50	*	.	.	-0.20	0.34
	Pro	201	.	.	B	.	.	T	.	-0.43	0.39	*	.	.	0.10	0.45
	Val	202	.	.	B	.	.	.	.	-0.06	0.27	.	.	F	0.03	0.39
	Pro	203	.	.	.	.	T	.	.	0.19	0.27	.	.	F	0.88	1.13
	Ser	204	.	.	.	.	T	.	.	0.80	0.13	.	.	F	1.01	0.72
	Ser	205	.	.	.	.	.	T	C	1.07	0.10	*	.	F	1.44	1.69
	Pro	206	.	.	.	.	T	T	.	0.47	-0.56	*	.	F	2.82	1.89
	Gly	207	.	.	.	.	T	T	.	0.66	-0.29	.	.	F	2.80	1.16
	Glu	208	.	.	S	.	.	T	.	0.52	-0.10	.	.	F	1.97	0.47
	Glu	209	.	.	S	.	.	.	.	0.82	-0.06	.	.	F	1.49	0.30
	Leu	210	.	.	S	.	.	.	.	1.12	-0.89	.	.	F	1.37	0.48
	Cys	211	.	.	S	.	.	T	.	1.33	-0.11	.	.	F	1.45	0.45
	Gly	212	.	.	.	.	T	T	.	0.82	-0.11	.	.	F	1.73	0.42
	Asn	213	.	.	.	.	T	T	.	0.51	0.53	.	*	F	0.99	0.38
	Asn	214	.	.	.	.	T	T	.	0.27	0.33	.	*	F	1.60	1.01
	Asn	215	.	.	.	B	T	.	.	0.19	0.51	.	.	F	0.74	1.60
	Val	216	.	.	B	B	.	.	.	0.56	0.77	*	.	.	-0.12	0.70
	Thr	217	.	.	B	B	.	.	.	0.69	0.76	*	.	.	-0.28	0.58
	Tyr	218	.	.	B	B	.	.	.	-0.07	0.74	.	.	.	-0.44	0.48
	Ile	219	.	.	B	B	.	.	.	-0.10	0.91	.	.	.	-0.60	0.35
	Ser	220	.	.	B	.	.	T	.	-0.70	0.77	*	*	.	-0.20	0.33
	Ser	221	.	.	B	.	.	T	.	0.27	0.90	*	.	.	-0.20	0.21
	Cys	222	.	.	B	.	.	T	.	0.58	0.14	*	.	.	0.10	0.58
	His	223	.	.	B	.	.	T	.	0.23	-0.14	*	.	.	0.70	0.75
	Met	224	.	.	.	.	T	.	.	0.81	-0.93	*	.	.	0.90	0.57

Table I (continued)

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Arg	225	.	.	B	B	.	.	.	0.44	0.07	*	*	.	-0.15	1.53
Gln	226	.	.	B	B	.	.	.	0.04	0.07	*	.	.	-0.30	0.60
Ala	227	.	.	B	B	.	.	.	-0.10	0.36	*	.	.	-0.30	0.53
Thr	228	.	.	B	B	.	.	.	-0.41	0.43	*	*	.	-0.60	0.22
Cys	229	.	.	B	B	.	.	.	0.30	0.86	*	*	.	-0.60	0.13
Phe	230	.	.	B	B	.	.	.	-0.11	0.46	*	.	.	-0.60	0.25
Leu	231	.	.	B	B	.	.	.	-1.08	0.34	*	.	.	-0.30	0.23
Gly	232	.	.	.	.	T	T	.	-0.76	0.54	*	.	.	0.20	0.30
Arg	233	.	.	.	.	T	T	.	-1.30	0.40	.	*	F	0.65	0.34
Ser	234	.	.	.	.	T	T	.	-0.52	0.26	.	.	F	0.65	0.31
Ile	235	.	.	B	.	.	T	.	0.14	-0.43	.	.	.	0.70	0.61
Gly	236	.	.	B	B	.	.	.	0.37	-0.36	.	.	.	0.30	0.42
Val	237	.	.	B	B	.	.	.	0.37	0.14	.	.	.	-0.30	0.32
Arg	238	.	.	B	B	.	.	.	-0.04	0.19	.	*	.	-0.30	0.45
His	239	.	.	B	.	.	T	.	-0.41	-0.11	*	*	.	0.70	0.63
Ala	240	.	.	.	.	T	T	.	-0.11	0.03	*	*	.	0.30	0.44
Gly	241	.	.	.	.	T	T	.	-0.11	-0.11	*	.	.	1.10	0.23
Ser	242	.	.	.	.	T	T	.	0.43	0.31	*	*	.	0.80	0.16
Cys	243	.	.	.	.	T	T	.	0.11	0.36	*	.	.	1.10	0.24
Ala	244	.	.	.	.	T	T	.	0.14	0.23	.	.	.	1.40	0.37
Gly	245	.	.	.	.	.	T	C	0.73	-0.20	.	.	F	2.25	0.47
Thr	246	.	.	.	.	.	T	C	0.87	-0.59	.	.	F	3.00	1.53
Pro	247	.	.	.	.	.	.	C	0.96	-0.73	.	.	F	2.30	2.35
Glu	248	.	.	.	.	.	.	C	1.28	-0.80	.	.	F	2.30	3.67
Glu	249	.	.	.	.	.	.	C	1.52	-0.80	.	.	F	2.50	2.52
Pro	250	.	.	.	.	.	T	C	1.87	-0.86	.	.	F	2.70	1.61
Pro	251	.	.	.	.	.	T	C	1.88	-1.29	.	.	F	2.70	1.61
Gly	252	.	.	.	.	.	T	C	1.50	-0.90	.	.	F	3.00	1.25
Gly	253	.	.	.	.	.	T	C	1.50	-0.40	.	.	F	2.25	0.81
Glu	254	.	A	.	.	.	.	C	1.50	-0.83	.	.	F	1.85	0.91
Ser	255	.	A	.	.	.	.	C	1.71	-1.26	.	.	F	1.70	1.60
Ala	256	A	A	.	.	.	.	.	1.92	-1.69	*	.	F	1.20	2.79
Glu	257	A	A	.	.	.	.	.	2.27	-2.11	*	.	F	0.90	2.79
Glu	258	A	A	.	.	.	.	.	1.91	-1.71	*	.	F	0.90	3.35
Glu	259	A	A	.	.	.	.	.	1.06	-1.31	*	.	F	0.90	2.87
Glu	260	A	A	.	.	.	.	.	0.97	-1.17	*	.	F	0.90	1.23
Asn	261	A	A	.	.	.	.	.	1.17	-0.74	*	.	.	0.60	0.91
Phe	262	A	A	.	.	.	.	.	0.78	-0.31	.	.	.	0.30	0.67
Val	263	A	A	.	.	.	.	.	0.39	0.11	.	.	.	-0.30	0.49

Among highly preferred fragments in this regard are those that comprise reigons of follistatin-3 that combine several structural features, such as several of the features set out above.

In another aspect, the invention provides an isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a portion of the polynucleotide in a nucleic acid molecule of the invention described above, for instance, the cDNA clone contained in ATCC Deposit No. 209199. By "stringent hybridization conditions" is intended overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

By a polynucleotide which hybridizes to a "portion" of a polynucleotide is intended a polynucleotide (either DNA or RNA) hybridizing to at least about 15 nucleotides (nt), and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably about 30-70 (e.g., 50) nt of the reference polynucleotide. These are useful as diagnostic probes and primers as discussed above and in more detail below.

By a portion of a polynucleotide of "at least 20 nt in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotide (e.g., the deposited cDNA or the nucleotide sequence as shown in Figures 1A, 1B, and 1C (SEQ ID NO:1)). Of course, a polynucleotide which hybridizes only to a poly A sequence (such as the 3' terminal poly(A) tract of the follistatin-3 cDNA shown in Figures 1A, 1B, and 1C (SEQ ID NO:1)), or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

In preferred embodiments, polynucleotides which hybridize to the reference polynucleotides disclosed herein encode polypeptides which either retain substantially the same biological function or activity as the mature form of the follistatin-3 polypeptide

encoded by the polynucleotide sequence depicted in Figures 1A, 1B, and 1C (SEQ ID NO:1) or the clone contained in the deposit (HDTAH85).

Alternative embodiments are directed to polynucleotides which hybridize to the reference polynucleotide (i.e., a polynucleotide sequence disclosed herein), but do not retain biological activity. While these polynucleotides do not retain biological activity, they have uses, such as, for example, as probes for the polynucleotides of SEQ ID NO:1, for recovery of the polynucleotides, as diagnostic probes, and as PCR primers.

As indicated, nucleic acid molecules of the present invention which encode a follistatin-3 polypeptide may include, but are not limited to those encoding the amino acid sequence of the mature polypeptide, by itself, and the coding sequence for the mature polypeptide and additional sequences, such as those encoding the about 26 amino acid leader or secretory sequence, such as a pre-, or pro- or prepro- protein sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences.

Also encoded by nucleic acids of the invention are the above protein sequences together with additional, non-coding sequences, including for example, but not limited to introns and non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription, mRNA processing, including splicing and polyadenylation signals, for example - ribosome binding and stability of mRNA; an additional coding sequence which codes for additional amino acids, such as those which provide additional functionalities.

Thus, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described by Gentz and colleagues (*Proc. Natl. Acad. Sci. USA* **86**:821-824 (1989)), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which

has been described by Wilson and coworkers (*Cell* 37:767 (1984)). As discussed below, other such fusion proteins include the follistatin-3 fused to Fc at the N- or C-terminus.

The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives of the follistatin-3 protein. Variants may occur naturally, such as a natural allelic variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism (*Genes II*, Lewin, B., ed., John Wiley & Sons, New York (1985)). Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

Such variants include those produced by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding regions, non-coding regions, or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the follistatin-3 protein or portions thereof. Also especially preferred in this regard are conservative substitutions.

Most highly preferred are nucleic acid molecules encoding the mature protein having the amino acid sequence shown in SEQ ID NO:2 or the mature follistatin-3 amino acid sequence encoded by the deposited cDNA clone.

Further embodiments include an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical to a polynucleotide selected from the group consisting of: (a) a nucleotide sequence encoding the follistatin-3 polypeptide having the complete amino acid sequence in SEQ ID NO:2 (i.e., positions -26 to 237 of SEQ ID NO:2); (b) a nucleotide sequence encoding the follistatin-3 polypeptide having the complete amino acid sequence in SEQ ID NO:2 excepting the N-terminal methionine (i.e., positions -25 to 237 of SEQ ID NO:2); (c) a nucleotide sequence encoding the predicted mature follistatin-3 polypeptide having the amino acid sequence at positions 1 to 237 in SEQ ID NO:2; (d) a nucleotide sequence encoding the follistatin-3 polypeptide having the complete amino acid sequence encoded by the cDNA clone

contained in ATCC Deposit No. 209199; (e) a nucleotide sequence encoding the follistatin-3 polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209199; (f) a nucleotide sequence encoding the mature follistatin-3 polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199; and (g) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), (d), (e) or (f) above.

Further embodiments of the invention include isolated nucleic acid molecules that comprise a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical, to any of the nucleotide sequences in (a), (b), (c), (d), (e), (f) or (g), above, or a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide in (a), (b), (c), (d), (e), (f) or (g), above. This polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues. An additional nucleic acid embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a follistatin-3 polypeptide having an amino acid sequence in (a), (b), (c), (d), (e) or (f), above. A further nucleic acid embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of a follistatin-3 polypeptide having an amino acid sequence which contains at least one conservative amino acid substitution, but not more than 50 conservative amino acid substitutions, even more preferably, not more than 40 conservative amino acid substitutions, still more preferably not more than 30 conservative amino acid substitutions, and still even more preferably not more than 20 conservative amino acid substitutions. Of course, in order of ever-increasing preference, it is highly preferable for a polynucleotide which encodes the amino acid sequence of a follistatin-3 polypeptide to have an amino acid sequence which contains not more than 7-10, 5-10, 3-7, 3-5, 2-5, 1-5, 1-3, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 conservative amino acid substitutions.

The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, and to host cells containing the

recombinant vectors, as well as to methods of making such vectors and host cells and for using them for production of follistatin-3 polypeptides or peptides by recombinant techniques.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a follistatin-3 polypeptide is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequences encoding the follistatin-3 polypeptides. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular nucleic acid molecule is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the nucleotide sequence shown in Figures 1A, 1B, and 1C, or to the nucleotides sequence of the deposited cDNA clone can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). Bestfit uses the local homology algorithm of Smith and Waterman to find the best segment of homology between two sequences (*Advances in Applied Mathematics* **2**:482-489 (1981)). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference nucleotide sequence and that gaps in homology of up to 5% of the total number of nucleotides in the reference sequence are allowed. A preferred method for determing the best overall

match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag and colleagues (*Comp. App. Biosci.* **6**:237-245 (1990)). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignement of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so

10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The present application is directed to nucleic acid molecules at least 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in Figures 1A, 1B, and 1C (SEQ ID NO:1) or to the nucleic acid sequence of the deposited cDNA, irrespective of whether they encode a polypeptide having follistatin-3 activity. This is because even where a particular nucleic acid molecule does not encode a polypeptide having follistatin-3 activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or a polymerase chain reaction (PCR) primer. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having follistatin-3 activity include, *inter alia*, (1) isolating the follistatin-3 gene or allelic variants thereof in a cDNA library; (2) *in situ* hybridization (e.g., "FISH") to metaphase chromosomal spreads to provide precise chromosomal location of the follistatin-3 gene, as described by Verma and colleagues (*Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York (1988)); and Northern Blot analysis for detecting follistatin-3 mRNA expression in specific tissues.

Preferred, however, are nucleic acid molecules having sequences at least 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in Figures 1A, 1B, and 1C (SEQ ID NO:1) or to the nucleic acid sequence of the deposited cDNA which do, in fact, encode a polypeptide having follistatin-3 protein activity. By "a polypeptide having follistatin-3 activity" is intended polypeptides exhibiting activity similar, but not necessarily identical, to an activity of the mature follistatin-3 protein of the invention, as measured in a particular biological assay. For example, the follistatin-3

protein of the present invention inhibits the binding of activin to the activin receptor. An activin receptor-binding inhibition assay is described by Hashimoto and colleagues (*J. Biol. Chem.* **272**:13835-13842 (1997)). Briefly, the assay involves culturing rat pituitary cells (5×10^5 cells) in 24-well plates in the presence of [^{125}I]-activin A (40 ng/mL; activin A is labeled using the chloramine-T method as described by Hasegawa and coworkers (*Endocrinol. Japan* **33**:645-654 (1986)) and follistatin-3 or a mutein thereof (200ng/mL). A baseline of activin-binding is determined by affinity cross-linking [^{125}I]-activin A to the pituitary cells using the bifunctional chemical cross-linker disuccinimidyl suberate (DSS) in the absence of follistatin-3. Cross-linking is achieved by washing cells once with binding buffer (DMEM containing 25 mM HEPES (pH 7.4) and 0.2% bovine serum albumen) and incubating on ice for 2 h with 40 ng/mL [^{125}I]-activin A in the binding buffer. Following incubation, cells are washed 3 times with ice-cold PBS and incubated in PBS containing 1 mM DSS for 20 min on ice. The reaction is then quenched with PBS. The cells are removed from the culture dish by scraping, rinsed with a Tris solution (20 mM Tris-HCl (pH 7.2) containing 2 mM EDTA, 5 mM benzamidine, 2 mM phenylmethylsulfonyl fluoride (PMSF), 2 mM *N*-ethylaleimide, and 2 mM diisopropyl fluorophosphate), centrifuged, and resuspended in solubilization buffer (50 mM Tris-HCl (pH 7.2) containing 150 mM NaCl, 2 mM EDTA, 5 mM benzamidine, 2 mM PMSF, 2 mM *N*-ethylaleimide, 2 mM diisopropyl fluorophosphate, 1% Triton X-100, and 10% glycerol), and stirred gently on ice for 1 h.The cell lysates are introduced into 2% SDS and boiled at 100°C for 10 min. The resulting affinity-labeled lysates are then subject to SDS-PAGE (7.5 or 8% gels). Following SDS-PAGE, gels are fixed, stained with 0.25% Coomassie Brilliant Blue R-250, destained, air-dried, and then visualized by autoradiography. Inhibition of activin binding of the activin receptor is analyzed in samples with which follistatin-3 or a mutein thereof (200ng/mL) are incubated with labeled activin in the binding buffer incubation described above. The degree to which the formation of affinity cross-linked activin/activin receptor complexes is decreased correlates with the ability of follistatin-3 or a mutein thereof to bind to labeled activin protein. As such, the relative binding affinity of activin for its receptor versus follistatin-3 or a mutein thereof can be

quantitated. Such activity is useful for regulating the effective amount of activin present in a given system.

Follistatin-3 protein binds to activin in a dose-dependent manner in the above-described assay. Thus, "a polypeptide having follistatin-3 protein activity" includes polypeptides that also exhibit any of the same binding activities in the above-described assays in a dose-dependent manner. Although the degree of dose-dependent activity need not be identical to that of the follistatin-3 protein, preferably, "a polypeptide having follistatin-3 protein activity" will exhibit substantially similar dose-dependence in a given activity as compared to the follistatin-3 protein (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity relative to the reference follistatin-3 protein).

Like follistatin-1, follistatin-3 inhibits the secretion of FSH. An assay for measuring the suppression of spontaneous FSH release from primary cultured rat pituitary cells is well known in the art (Hasegawa, Y., *et al., Endocrinol. Jpn.* **33**:645-654 (1986)). Briefly, freshly isolated pituitary cells are suspended in DMEM containing gentamicin (35 μg/mL), fungizone (1 μg/mL), 0.05% glutamine, 0.1% sodium bicarbonate, 10% horse serum, and 2.5% fetal bovine serum at a density of 3 x 10^5 cells/mL, and plated in 96-well culture plates (6 x 10^5 cells/0.2 mL/well). Various amounts (0.1-100 ng/mL) of follistatin-3 are then added to the culture medium. After culturing for 3 days at 37°C (5% CO_2), cultured media are assayed for quantity of secreted FSH by a double antibody RIA method using an RIA kit and plotted as FSH Secreted (ng/mL/72 h) versus Protein Added (ng/mL).

Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the nucleic acid molecules having a sequence at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the nucleic acid sequence of the deposited cDNA or the nucleic acid sequence shown in Figures 1A, 1B, and 1C (SEQ ID NO:1) will encode a polypeptide "having follistatin-3 protein activity." In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a

polypeptide having follistatin-3 protein activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid), as further described below.

The present invention also relates to vectors which include the isolated DNA molecules of the present invention, host cells which are genetically engineered with the recombinant vectors, and the production of follistatin-3 polypeptides or fragments thereof by recombinant techniques. The vector may be, for example, a phage, plasmid, viral or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli lac*, *trp*, *phoA* and *tac* promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as

Drosophila S2 and Spodoptera Sf9 cells, animal cells such as CHO, COS, 293 and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9 (QIAGEN, Inc., *supra*); pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1, and pSG (Stratagene); and pSVK3, pBPV, pMSG and pSVL (Pharmacia). Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals (for example, Davis, *et al.*, *Basic Methods In Molecular Biology* (1986)).

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous functional regions. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to stabilize and purify proteins. For example, EP-A-O 464 533 (Canadian counterpart 20458869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is thoroughly advantageous for use in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been

expressed, detected and purified in the advantageous manner described. This is the case when Fc portion proves to be a hindrance to use in therapy and diagnosis, for example when the fusion protein is to be used as antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5 (Bennett, D., *et al., J. Molecular Recognition* **8**:52-58 (1995); Johanson, K., *et al., J. Biol. Chem.* **270**:9459-9471 (1995)).

The follistatin-3 protein can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. Polypeptides of the present invention include: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Polypeptides and Fragments

The invention further provides an isolated follistatin-3 polypeptide having the amino acid sequence encoded by the deposited cDNA, or the amino acid sequence in

SEQ ID NO:2, or a peptide or polypeptide comprising a portion of the above polypeptides.

To improve or alter the characteristics of follistatin-3 polypeptides, protein engineering may be employed. Recombinant DNA technology known to those skilled in the art can be used to create novel mutant proteins or muteins including single or multiple amino acid substitutions, deletions, additions or fusion proteins. Such modified polypeptides can show, e.g., enhanced activity or increased stability. In addition, they may be purified in higher yields and show better solubility than the corresponding natural polypeptide, at least under certain purification and storage conditions.

For instance, for many proteins, including the extracellular domain of a membrane associated protein or the mature form(s) of a secreted protein, it is known in the art that one or more amino acids may be deleted from the N-terminus or C-terminus without substantial loss of biological function. For instance, Ron and colleagues (*J. Biol. Chem.*, **268**:2984-2988 (1993)) reported modified KGF proteins that had heparin binding activity even if 3, 8, or 27 N-terminal amino acid residues were missing. In the present case, since the protein of the invention is a member of the inhibin-related polypeptide family, deletions of N-terminal amino acids up to the cysteine at position 12 of SEQ ID NO:2 may retain some biological activity such as binding activin or an activin-like molecule. Polypeptides having further N-terminal deletions including the cysteine-12 residue in SEQ ID NO:2 would not be expected to retain such biological activities because it is known that this residue is likely required for forming a disulfide bridge to provide structural stability which is needed for protein-protein interaction and is in the beginning of the conserved domain required for biological activities.

However, even if deletion of one or more amino acids from the N-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened protein to induce and/or bind to antibodies which recognize the complete or mature of the protein generally will be retained when less than the majority of the residues of the complete or mature protein are removed from the N-terminus. Whether a particular polypeptide lacking N-terminal residues of a complete protein retains such immunologic

activities can readily be determined by routine methods described herein and otherwise known in the art.

Accordingly, the present invention further provides polypeptides having one or more residues deleted from the amino terminus of the amino acid sequence of the follistatin-3 shown in SEQ ID NO:2, up to the cysteine residue at position number 12, and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides comprising the amino acid sequence of residues n^1-237 of SEQ ID NO:2, where n^1 is an integer in the range of -26-12, and 12 is the position of the first residue from the N-terminus of the complete follistatin-3 polypeptide (shown in SEQ ID NO:2) believed to be required for activin-binding or activin-like protein-binding activity of the follistatin-3 protein.

More in particular, the invention provides polynucleotides encoding polypeptides having the amino acid sequence of residues of -26-237, -25-237, -24-237, -23-237, -22-237, -21-237, -20-237, -19-237, -18-237, -17-237, -16-237, -15-237, -14-237, -13-237, -12-237, -10-237, -9-237, -8-237, -7-237, -6-237, -5-237, -4-237, -3-237, -2-237, -1-237, 1-237, 2-237, 3-237, 4-237, 5-237, 6-237, 7-237, 8-237, 9-237, 10-237, 11-237, and 12-237 of SEQ ID NO:2. Polypeptides encoded by these polynucleotides also are provided. The present invention is also directed to nucleic acid molecules comprising, or alternatively, consisting of, a polynucleotide sequence at least 80%, 85%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99% identical to the polynucleotide sequences encoding the follistatin-3 polypeptides described above, and the polypeptides encoded thereby. The present invention also encompasses the above polynucleotide sequences fused to a heterologous polynucleotide sequence, and the polypeptides encoded thereby.

Similarly, many examples of biologically functional C-terminal deletion muteins are known. For instance, Interferon gamma shows up to ten times higher activities by deleting 8-10 amino acid residues from the carboxy terminus of the protein (Dobeli, *et al., J. Biotechnology* 7:199-216 (1988)). In the present case, since the protein of the invention is a member of the activin-related polypeptide family, deletions of C-terminal amino acids up to the cysteine at position 217 of SEQ ID NO:2 may retain some biological activity such as binding activin or an activin-like molecule. Polypeptides having further C-terminal deletions including the cysteine residue at position 217 of SEQ

ID NO:2 would not be expected to retain such biological activities because it is known that this residue is likely required for forming a disulfide bridge to provide structural stability which is needed for protein-protein interactions and is the beginning of the conserved domain required for biological activities.

However, even if deletion of one or more amino acids from the C-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened protein to induce and/or bind to antibodies which recognize the complete or mature form of the protein generally will be retained when less than the majority of the residues of the complete or mature form of the protein are removed from the C-terminus. Whether a particular polypeptide lacking C-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art.

Accordingly, the present invention further provides polypeptides having one or more residues from the carboxy terminus of the amino acid sequence of the follistatin-3 shown in SEQ ID NO:2, up to the cysteine residue at position 217 of SEQ ID NO:2, and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides having the amino acid sequence of residues -26-m^1 of the amino acid sequence in SEQ ID NO:2, where m^1 is any integer in the range of 217 to 237, and residue 217 is the position of the first residue from the C- terminus of the complete follistatin-3 polypeptide (shown in SEQ ID NO:2) believed to be required for the activin-binding or activin-like protein-binding of the follistatin-3 protein.

More in particular, the invention provides polynucleotides encoding polypeptides having the amino acid sequence of residues -26-217, -26-218, -26-219, -26-220, -26-221, -26-222, -26-223, -26-224, -26-225, -26-226, -26-227, -26-228, -26-229, -26-230, -26-231, -26-232, -26-233, -26-234, -26-235, -26-236, and -26-237 of SEQ ID NO:2. Polypeptides encoded by these polynucleotides also are provided. The present invention is also directed to nucleic acid molecules comprising, or alternatively, consisting of, a polynucleotide sequence at least 80%, 85%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99% identical to the polynucleotide sequences encoding the follistatin-3 polypeptides described above, and the polypeptides encoded thereby. The

present invention also encompasses the above polynucleotide sequences fused to a heterologous polynucleotide sequence, and the polypeptides encoded thereby.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues n^1-m^1 of SEQ ID NO:2, where n^1 and m^1 are integers as described above.

Also included are a nucleotide sequence encoding a polypeptide consisting of a portion of the complete follistatin-3 amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199, where this portion excludes from 1 to about 37 amino acids from the amino terminus of the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199, or from 1 to about 20 amino acids from the carboxy terminus, or any combination of the above amino terminal and carboxy terminal deletions, of the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199. Polynucleotides encoding all of the above deletion mutant polypeptide forms also are provided.

As mentioned above, even if deletion of one or more amino acids from the N-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened follistatin-3 mutein to induce and/or bind to antibodies which recognize the complete or mature of the protein generally will be retained when less than the majority of the residues of the complete or mature protein are removed from the N-terminus. Whether a particular polypeptide lacking N-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a follistatin-3 mutein with a large number of deleted N-terminal amino acid residues may retain some biological or immungenic activities. In fact, peptides composed of as few as six follistatin-3 amino acid residues may often evoke an immune response.

Accordingly, the present invention further provides polypeptides having one or more residues deleted from the amino terminus of the follistatin-3 amino acid sequence shown in SEQ ID NO:2, up to the glutamic acid residue at position number 258 and polynucleotides encoding such polypeptides. In particular, the present invention

provides polypeptides comprising the amino acid sequence of residues n^2-263 of Figure 1A (SEQ ID NO:2), where n^2 is an integer in the range of 2 to 258, and 259 is the position of the first residue from the N-terminus of the complete follistatin-3 polypeptide believed to be required for at least immunogenic activity of the follistatin-3 protein.

More in particular, the invention provides polynucleotides encoding polypeptides comprising, or alternatively consisting of, the amino acid sequence of residues of R-2 to V-263; P-3 to V-263; G-4 to V-263; A-5 to V-263; P-6 to V-263; G-7 to V-263; P-8 to V-263; L-9 to V-263; W-10 to V-263; P-11 to V-263; L-12 to V-263; P-13 to V-263; W-14 to V-263; G-15 to V-263; A-16 to V-263; L-17 to V-263; A-18 to V-263; W-19 to V-263; A-20 to V-263; V-21 to V-263; G-22 to V-263; F-23 to V-263; V-24 to V-263; S-25 to V-263; S-26 to V-263; M-27 to V-263; G-28 to V-263; S-29 to V-263; G-30 to V-263; N-31 to V-263; P-32 to V-263; A-33 to V-263; P-34 to V-263; G-35 to V-263; G-36 to V-263; V-37 to V-263; C-38 to V-263; W-39 to V-263; L-40 to V-263; Q-41 to V-263; Q-42 to V-263; G-43 to V-263; Q-44 to V-263; E-45 to V-263; A-46 to V-263; T-47 to V-263; C-48 to V-263; S-49 to V-263; L-50 to V-263; V-51 to V-263; L-52 to V-263; Q-53 to V-263; T-54 to V-263; D-55 to V-263; V-56 to V-263; T-57 to V-263; R-58 to V-263; A-59 to V-263; E-60 to V-263; C-61 to V-263; C-62 to V-263; A-63 to V-263; S-64 to V-263; G-65 to V-263; N-66 to V-263; I-67 to V-263; D-68 to V-263; T-69 to V-263; A-70 to V-263; W-71 to V-263; S-72 to V-263; N-73 to V-263; L-74 to V-263; T-75 to V-263; H-76 to V-263; P-77 to V-263; G-78 to V-263; N-79 to V-263; K-80 to V-263; I-81 to V-263; N-82 to V-263; L-83 to V-263; L-84 to V-263; G-85 to V-263; F-86 to V-263; L-87 to V-263; G-88 to V-263; L-89 to V-263; V-90 to V-263; H-91 to V-263; C-92 to V-263; L-93 to V-263; P-94 to V-263; C-95 to V-263; K-96 to V-263; D-97 to V-263; S-98 to V-263; C-99 to V-263; D-100 to V-263; G-101 to V-263; V-102 to V-263; E-103 to V-263; C-104 to V-263; G-105 to V-263; P-106 to V-263; G-107 to V-263; K-108 to V-263; A-109 to V-263; C-110 to V-263; R-111 to V-263; M-112 to V-263; L-113 to V-263; G-114 to V-263; G-115 to V-263; R-116 to V-263; P-117 to V-263; R-118 to V-263; C-119 to V-263; E-120 to V-263; C-121 to V-263; A-122 to V-263; P-123 to V-263; D-124 to V-263; C-125 to V-263; S-126 to V-263; G-127 to V-263; L-128 to V-263; P-129 to V-263; A-130 to V-263; R-131 to V-263; L-132 to V-263; Q-133 to V-263; V-134 to V-263; C-135 to

V-263; G-136 to V-263; S-137 to V-263; D-138 to V-263; G-139 to V-263; A-140 to V-263; T-141 to V-263; Y-142 to V-263; R-143 to V-263; D-144 to V-263; E-145 to V-263; C-146 to V-263; E-147 to V-263; L-148 to V-263; R-149 to V-263; A-150 to V-263; A-151 to V-263; R-152 to V-263; C-153 to V-263; R-154 to V-263; G-155 to V-263; H-156 to V-263; P-157 to V-263; D-158 to V-263; L-159 to V-263; S-160 to V-263; V-161 to V-263; M-162 to V-263; Y-163 to V-263; R-164 to V-263; G-165 to V-263; R-166 to V-263; C-167 to V-263; R-168 to V-263; K-169 to V-263; S-170 to V-263; C-171 to V-263; E-172 to V-263; H-173 to V-263; V-174 to V-263; V-175 to V-263; C-176 to V-263; P-177 to V-263; R-178 to V-263; P-179 to V-263; Q-180 to V-263; S-181 to V-263; C-182 to V-263; V-183 to V-263; V-184 to V-263; D-185 to V-263; Q-186 to V-263; T-187 to V-263; G-188 to V-263; S-189 to V-263; A-190 to V-263; H-191 to V-263; C-192 to V-263; V-193 to V-263; V-194 to V-263; C-195 to V-263; R-196 to V-263; A-197 to V-263; A-198 to V-263; P-199 to V-263; C-200 to V-263; P-201 to V-263; V-202 to V-263; P-203 to V-263; S-204 to V-263; S-205 to V-263; P-206 to V-263; G-207 to V-263; Q-208 to V-263; E-209 to V-263; L-210 to V-263; C-211 to V-263; G-212 to V-263; N-213 to V-263; N-214 to V-263; N-215 to V-263; V-216 to V-263; T-217 to V-263; Y-218 to V-263; I-219 to V-263; S-220 to V-263; S-221 to V-263; C-222 to V-263; H-223 to V-263; M-224 to V-263; R-225 to V-263; Q-226 to V-263; A-227 to V-263; T-228 to V-263; C-229 to V-263; F-230 to V-263; L-231 to V-263; G-232 to V-263; R-233 to V-263; S-234 to V-263; I-235 to V-263; G-236 to V-263; V-237 to V-263; R-238 to V-263; H-239 to V-263; A-240 to V-263; G-241 to V-263; S-242 to V-263; C-243 to V-263; A-244 to V-263; G-245 to V-263; T-246 to V-263; P-247 to V-263; E-248 to V-263; E-249 to V-263; P-250 to V-263; P-251 to V-263; G-252 to V-263; G-253 to V-263; E-254 to V-263; S-255 to V-263; A-256 to V-263; E-257 to V-263; and E-258 to V-263 of the follistatin-3 amino acid sequence shown in Figure 1A (which is identical to the sequence shown as SEQ ID NO:2, with the exception that the amino acid residues in Figure 1A are numbered consecutively from 1 through 263 from the N-terminus to the C-terminus, while the amino acid residues in SEQ ID NO:2 are numbered consecutively from -26 through 237 to reflect the position of the predicted signal peptide). Polypeptides encoded by these polynucleotides also are provided. The present invention is also directed to nucleic acid

molecules comprising, or alternatively, consisting of, a polynucleotide sequence at least 80%, 85%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99% identical to the polynucleotide sequences encoding the follistatin-3 polypeptides described above, and the polypeptides encoded thereby. The present invention also encompasses the above polynucleotide sequences fused to a heterologous polynucleotide sequence, and the polypeptides encoded thereby.

Also as mentioned above, even if deletion of one or more amino acids from the C-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened follistatin-3 mutein to induce and/or bind to antibodies which recognize the complete or mature of the protein generally will be retained when less than the majority of the residues of the complete or mature protein are removed from the C-terminus. Whether a particular polypeptide lacking C-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a follistatin-3 mutein with a large number of deleted C-terminal amino acid residues may retain some biological or immungenic activities. In fact, peptides composed of as few as six follistatin-3 amino acid residues may often evoke an immune response.

Accordingly, the present invention further provides polypeptides having one or more residues deleted from the carboxy terminus of the amino acid sequence of the follistatin-3 shown in SEQ ID NO:2, up to the proline residue at position number 6, and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides comprising the amino acid sequence of residues 1-m^2 of SEQ ID NO:2, where m^2 is an integer in the range of 6 to 262, and 6 is the position of the first residue from the C-terminus of the complete follistatin-3 polypeptide believed to be required for at least immunogenic activity of the follistatin-3 protein.

More in particular, the invention provides polynucleotides encoding polypeptides comprising, or alternatively consisting of, the amino acid sequence of residues M-1 to F-262; M-1 to N-261; M-1 to E-260; M-1 to E-259; M-1 to E-258; M-1 to E-257; M-1 to A-256; M-1 to S-255; M-1 to E-254; M-1 to G-253; M-1 to G-252; M-1 to P-251; M-1 to P-250; M-1 to E-249; M-1 to E-248; M-1 to P-247; M-1 to T-246; M-1 to

G-245; M-1 to A-244; M-1 to C-243; M-1 to S-242; M-1 to G-241; M-1 to A-240; M-1 to H-239; M-1 to R-238; M-1 to V-237; M-1 to G-236; M-1 to I-235; M-1 to S-234; M-1 to R-233; M-1 to G-232; M-1 to L-231; M-1 to F-230; M-1 to C-229; M-1 to T-228; M-1 to A-227; M-1 to Q-226; M-1 to R-225; M-1 to M-224; M-1 to H-223; M-1 to C-222; M-1 to S-221; M-1 to S-220; M-1 to I-219; M-1 to Y-218; M-1 to T-217; M-1 to V-216; M-1 to N-215; M-1 to N-214; M-1 to N-213; M-1 to G-212; M-1 to C-211; M-1 to L-210; M-1 to E-209; M-1 to Q-208; M-1 to G-207; M-1 to P-206; M-1 to S-205; M-1 to S-204; M-1 to P-203; M-1 to V-202; M-1 to P-201; M-1 to C-200; M-1 to P-199; M-1 to A-198; M-1 to A-197; M-1 to R-196; M-1 to C-195; M-1 to V-194; M-1 to V-193; M-1 to C-192; M-1 to H-191; M-1 to A-190; M-1 to S-189; M-1 to G-188; M-1 to T-187; M-1 to Q-186; M-1 to D-185; M-1 to V-184; M-1 to V-183; M-1 to C-182; M-1 to S-181; M-1 to Q-180; M-1 to P-179; M-1 to R-178; M-1 to P-177; M-1 to C-176; M-1 to V-175; M-1 to V-174; M-1 to H-173; M-1 to E-172; M-1 to C-171; M-1 to S-170; M-1 to K-169; M-1 to R-168; M-1 to C-167; M-1 to R-166; M-1 to G-165; M-1 to R-164; M-1 to Y-163; M-1 to M-162; M-1 to V-161; M-1 to S-160; M-1 to L-159; M-1 to D-158; M-1 to P-157; M-1 to H-156; M-1 to G-155; M-1 to R-154; M-1 to C-153; M-1 to R-152; M-1 to A-151; M-1 to A-150; M-1 to R-149; M-1 to L-148; M-1 to E-147; M-1 to C-146; M-1 to E-145; M-1 to D-144; M-1 to R-143; M-1 to Y-142; M-1 to T-141; M-1 to A-140; M-1 to G-139; M-1 to D-138; M-1 to S-137; M-1 to G-136; M-1 to C-135; M-1 to V-134; M-1 to Q-133; M-1 to L-132; M-1 to R-131; M-1 to A-130; M-1 to P-129; M-1 to L-128; M-1 to G-127; M-1 to S-126; M-1 to C-125; M-1 to D-124; M-1 to P-123; M-1 to A-122; M-1 to C-121; M-1 to E-120; M-1 to C-119; M-1 to R-118; M-1 to P-117; M-1 to R-116; M-1 to G-115; M-1 to G-114; M-1 to L-113; M-1 to M-112; M-1 to R-111; M-1 to C-110; M-1 to A-109; M-1 to K-108; M-1 to G-107; M-1 to P-106; M-1 to G-105; M-1 to C-104; M-1 to E-103; M-1 to V-102; M-1 to G-101; M-1 to D-100; M-1 to C-99; M-1 to S-98; M-1 to D-97; M-1 to K-96; M-1 to C-95; M-1 to P-94; M-1 to L-93; M-1 to C-92; M-1 to H-91; M-1 to V-90; M-1 to L-89; M-1 to G-88; M-1 to L-87; M-1 to F-86; M-1 to G-85; M-1 to L-84; M-1 to L-83; M-1 to N-82; M-1 to I-81; M-1 to K-80; M-1 to N-79; M-1 to G-78; M-1 to P-77; M-1 to H-76; M-1 to T-75; M-1 to L-74; M-1 to N-73; M-1 to S-72; M-1 to W-71; M-1 to A-70; M-1 to T-69; M-1 to D-68; M-1 to

I-67; M-1 to N-66; M-1 to G-65; M-1 to S-64; M-1 to A-63; M-1 to C-62; M-1 to C-61; M-1 to E-60; M-1 to A-59; M-1 to R-58; M-1 to T-57; M-1 to V-56; M-1 to D-55; M-1 to T-54; M-1 to Q-53; M-1 to L-52; M-1 to V-51; M-1 to L-50; M-1 to S-49; M-1 to C-48; M-1 to T-47; M-1 to A-46; M-1 to E-45; M-1 to Q-44; M-1 to G-43; M-1 to Q-42; M-1 to Q-41; M-1 to L-40; M-1 to W-39; M-1 to C-38; M-1 to V-37; M-1 to G-36; M-1 to G-35; M-1 to P-34; M-1 to A-33; M-1 to P-32; M-1 to N-31; M-1 to G-30; M-1 to S-29; M-1 to G-28; M-1 to M-27; M-1 to S-26; M-1 to S-25; M-1 to V-24; M-1 to F-23; M-1 to G-22; M-1 to V-21; M-1 to A-20; M-1 to W-19; M-1 to A-18; M-1 to L-17; M-1 to A-16; M-1 to G-15; M-1 to W-14; M-1 to P-13; M-1 to L-12; M-1 to P-11; M-1 to W-10; M-1 to L-9; M-1 to P-8; M-1 to G-7; M-1 to P-6 of the sequence of the follistatin-3 sequence shown in Figure 1A (which is identical to the sequence shown as SEQ ID NO:2, with the exception that the amino acid residues in Figure 1A are numbered consecutively from 1 through 263 from the N-terminus to the C-terminus, while the amino acid residues in SEQ ID NO:2 are numbered consecutively from -26 through 237 to reflect the position of the predicted signal peptide). Polypeptides encoded by these polynucleotides also are provided. The present invention is also directed to nucleic acid molecules comprising, or alternatively, consisting of, a polynucleotide sequence at least 80%, 85%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99% identical to the polynucleotide sequences encoding the follistatin-3 polypeptides described above, and the polypeptides encoded thereby. The present invention also encompasses the above polynucleotide sequences fused to a heterologous polynucleotide sequence, and the polypeptides encoded thereby.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini of a follistatin-3 polypeptide, which may be described generally as having residues n^2-m^2 of Figure 1A (SEQ ID NO:2), where n^2 and m^2 are integers as described above.

In addition to terminal deletion forms of the protein discussed above, it also will be recognized by one of ordinary skill in the art that some amino acid sequences of the follistatin-3 polypeptide can be varied without significant effect of the structure or function of the protein. If such differences in sequence are contemplated, it should be remembered that there will be critical areas on the protein which determine activity.

Thus, the invention further includes variations of the follistatin-3 polypeptide which show substantial follistatin-3 polypeptide activity or which include regions of follistatin-3 protein such as the protein portions discussed below. Such mutants include deletions, insertions, inversions, repeats, and type substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided wherein the authors indicate that there are two main approaches for studying the tolerance of an amino acid sequence to change (Bowie, J. U., *et al.*, *Science* **247**:1306-1310 (1990)). The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selections or screens to identify sequences that maintain functionality.

As the authors state, these studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at a certain position of the protein. For example, most buried amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described by Bowie and coworkers (*supra*) and the references cited therein. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Thus, the fragment, derivative or analog of the polypeptide of SEQ ID NO:2, or that encoded by the deposited cDNA, may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the

additional amino acids are fused to the above form of the polypeptide, such as an IgG Fc fusion region peptide or leader or secretory sequence or a sequence which is employed for purification of the above form of the polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Thus, the follistatin-3 of the present invention may include one or more amino acid substitutions, deletions or additions, either from natural mutations or human manipulation. As indicated, changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein (see Table II).

TABLE II. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine Tryptophan Tyrosine
Hydrophobic	Leucine Isoleucine Valine
Polar	Glutamine Asparagine
Basic	Arginine Lysine Histidine
Acidic	Aspartic Acid Glutamic Acid
Small	Alanine Serine Threonine Methionine Glycine

Embodiments of the invention are directed to polypeptides which comprise the amino acid sequence of a follistatin-3 polypeptide described hereub, but having an amino acid sequence which contains at least one conservative amino acid substitution, but not

more than 50 conservative amino acid substitutions, even more preferably, not more than 40 conservative amino acid substitutions, still more preferably, not more than 30 conservative amino acid substitutions, and still even more preferably, not more than 20 conservative amino acid substitutions, when compared with the follistatin-3 polynucleotide sequence described herein. Of course, in order of ever-increasing preference, it is highly preferable for a peptide or polypepside to have an amino acid sequence which comprises the amino acid sequence of a follistatin-3 polypeptide, which contains at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 conservative amino acid substitutions.

In further specific embodiments, the number of substitutions, additions or deletions in the amino acid sequence of Figures 1A, 1B, and 1C (SEQ ID NO:2), a polypeptide sequence encoded by the deposited clones, and/or any of the polypeptide fragments described herein is 75, 70, 60, 50, 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1 or 150-50, 100-50, 50-20, 30-20, 20-15, 20-10, 15-10, 10-1, 5-10, 1-5, 1-3 or 1-2.

To improve or alter the characteristics of follistatin-3 polypeptides, protein engineering may be employed. Recombinant DNA technology known to those skilled in the art can be used to create novel mutant proteins or muteins including single or multiple amino acid substitutions, deletions, additions or fusion proteins. Such modified polypeptides can show, e.g., enhanced activity or increased stability. In addition, they may be purified in higher yields and show better solubility than the corresponding natural polypeptide, at least under certain purification and storage conditions.

Thus, the invention also encompasses follistatin-3 derivatives and analogs that have one or more amino acid residues deleted, added, or substituted to generate follistatin-3 polypeptides that are better suited for expression, scale up, etc., in the host cells chosen. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges; N-linked glycosylation sites can be altered or eliminated to acheive, for example, expression of a homogeneous product that is more easily recovered and purified from yeast hosts which are known to hyperglycosylate N-linked sites. To this end, a variety of amino acid substitutions at one or both of the first or third amino acid positions on any one or more of the glycosylation recognitions sequences in the follistatin-3 polypeptides of the invention, and/or an amino

acid deletion at the second position of any one or more such recognition sequences will prevent glycosylation of the follistatin-3 polypeptide at the modified tripeptide sequence (see, e.g., Miyajima, A., *et al., EMBO J.* 5(6):1193-1197 (1986)).

Amino acids in the follistatin-3 protein of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science* **244:**1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro* proliferative activity.

Of special interest are substitutions of charged amino acids with other charged or neutral amino acids which may produce proteins with highly desirable improved characteristics, such as less aggregation. Aggregation may not only reduce activity but also be problematic when preparing pharmaceutical formulations, because aggregates can be immunogenic (Pinckard, *et al., Clin. Exp. Immunol.* **2:**331-340 (1967); Robbins, *et al., Diabetes* **36:**838-845 (1987); Cleland, *et al., Crit. Rev. Therapeutic Drug Carrier Systems* **10:**307-377 (1993)).

A mutational analysis of the two N-linked glycosylation sites (Asn-95 and Asn-259) of follistatin-1 was conducted by Inouye and colleagues (*Biochem. Biophys. Res. Comm.* **179:**352-358 (1991)). As described in the analysis, disruption of either or both of the N-linked glycosylation sites (by mutation of Thr-97 and Thr-261 to alanine) had no discernable effect on activin-binding and FSH secretion. However, results of the same study suggest that insertion of two amino acid residues (lysine and leucine) between residues Asn-2 and Cys-3 of follistatin-1 completely abolishes its inhibitory activity on FSH secretion from the pituitary, as well as its ability to bind activin. The asparagine and surrounding residues described in this analysis are weakly conserved between follistatin-1 and follistatin-3. There are however, two potential N-linked glycosylation sites in the sequence of follistatin-3 (N-73 and N-215; see Figure 1A). In addition, 4 out of 5 amino acids making up the sequence near the amino terminus, at which point Inouye and coworkers made their two amino acid insertion (*supra*), are conserved. Consequently, the extreme amino terminal region of the predicted mature

follistatin-3 polypeptide may have a high potential for exhibiting a deleterious effect through mutation.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of the follistatin-3 polypeptide can be substantially purified by the one-step method described by Smith and Johnson (*Gene* 67:31-40 (1988)). Polypeptides of the invention also can be purified from natural or recombinant sources using anti-Follistatin-3 antibodies of the invention in methods which are well known in the art of protein purification.

The invention further provides an isolated follistatin-3 polypeptide comprising an amino acid sequence selected from the group consisting of: (a) the amino acid sequence of the full-length follistatin-3 polypeptide having the complete amino acid sequence shown in SEQ ID NO:2 (i.e., positions -26 to 237 of SEQ ID NO:2); (b) the amino acid sequence of the full-length follistatin-3 polypeptide having the complete amino acid sequence shown in SEQ ID NO:2 excepting the N-terminal methionine (i.e., positions -25 to 237 of SEQ ID NO:2); (c) the amino acid sequence of the predicted mature follistatin-3 polypeptide having the amino acid sequence at positions 1 to 237 in SEQ ID NO:2; (d) the amino acid sequence of the full-length follistatin-3 polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199; (e) the amino acid sequence of the full-length follistatin-3 polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209199; and (f) the amino acid sequence of the mature follistatin-3 polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209199. The polypeptides of the present invention also include polypeptides having an amino acid sequence at least 80% identical, more preferably at least 90% identical, and still more preferably 95%, 96%, 97%, 98% or 99% identical to those described in (a), (b), (c), (d), (e) or (f) above, as well as polypeptides having an amino acid sequence with at least 90% similarity, and more preferably at least 95% similarity, to those above.

Further polypeptides of the present invention include polypeptides which have at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98% or 99% similarity to those described above. The polypeptides of

the invention also comprise those which are at least 80% identical, more preferably at least 90% or 95% identical, still more preferably at least 96%, 97%, 98% or 99% identical to the polypeptide encoded by the deposited cDNA or to the polypeptide of SEQ ID NO:2, and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

By "% similarity" for two polypeptides is intended a similarity score produced by comparing the amino acid sequences of the two polypeptides using the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) and the default settings for determining similarity. Bestfit uses the local homology algorithm of Smith and Waterman (Advances in Applied Mathematics 2:482-489, 1981) to find the best segment of similarity between two sequences.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a reference amino acid sequence of a follistatin-3 polypeptide is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of the follistatin-3 polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequence shown in Figure 1A (SEQ ID NO:2), the amino acid sequence encoded by deposited cDNA clone HDTAH85, or fragments thereof, can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package,

Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference amino acid sequence and that gaps in homology of up to 5% of the total number of amino acid residues in the reference sequence are allowed.

In a specific embodiment, the identity between a reference (query) sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, is determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. 6:237-245 (1990)). Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter. According to this embodiment, if the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction is made to the results to take into consideration the fact that the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. A determination of whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of this embodiment. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions

outside the farthest N- and C-terminal residues of the subject sequence. For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are made for the purposes of this embodiment.

The invention also encompasses fusion proteins in which the full-length follistatin-3 polypeptide or fragment, variant, derivative, or analog thereof is fused to an unrelated protein. These fusion proteins can be routinely designed on the basis of the follistatin-3 nucleotide and polypeptide sequences disclosed herein. For example, as one of skill in the art will appreciate, follistatin-3 polypeptides and fragments (including epitope-bearing fragments) thereof described herein can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric (fusion) polypeptides. These fusion proteins facilitate purification and show an increased half-life *in vivo*. This has been shown, e.g., for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (EP A 394,827; Traunecker, *et al., Nature* **331**:84-86 (1988)). Fusion proteins that have a disulfide-linked dimeric structure due to the IgG part can also be more efficient in binding and neutralizing other molecules than the monomeric follistatin-3 polypeptide or polypeptide fragments alone (Fountoulakis, *et*

al., J. Biochem. **270**:3958-3964 (1995)). Examples of follistatin-3 fusion proteins that are encompassed by the invention include, but are not limited to, fusion of the follistatin-3 polypeptide sequences to any amino acid sequence that allows the fusion proteins to be displayed on the cell surface (e.g. the IgG Fc domain); or fusions to an enzyme, fluorescent protein, or luminescent protein which provides a marker function.

The polypeptide of the present invention could be used as a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art.

As described in detail below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting follistatin-3 protein expression as described below or as agonists and antagonists capable of enhancing or inhibiting follistatin-3 protein function. Further, such polypeptides can be used in the yeast two-hybrid system to "capture" follistatin-3 protein binding proteins which are also candidate agonists and antagonists according to the present invention. The yeast two hybrid system is described by Fields and Song (*Nature* **340**:245-246 (1989)).

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention. The epitope of this polypeptide portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope". The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes (see, for instance, Geysen, *et al., Proc. Natl. Acad. Sci. USA* **81**:3998-4002 (1983)).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein (see, for instance, Sutcliffe, J. G., *et al., Science* **219**:660-666 (1983)). Peptides capable of eliciting protein-reactive sera are frequently represented in the

primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention (see, for instance, Wilson, *et al.*, *Cell* **37**:767-778 (1984)).

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate follistatin-3-specific antibodies include: a polypeptide comprising amino acid residues from about Leu-14 to about Ala-20, from about Ser-46 to about Ile-55, from about Gly-88 to about Pro-97, from about Gly-113 to about Leu-133, from about Arg-138 to about Glu-146, from about Pro-177 to about Thr-191, from about Gly-219 to about Val-237. These polypeptide fragments have been determined to bear antigenic epitopes of the follistatin-3 protein by the analysis of the Jameson-Wolf antigenic index, as shown in Figure 3 and Table I, above.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means (see, for example, Houghten, R. A., *et al.*, *Proc. Natl. Acad. Sci. USA* 82:5131-5135 (1985); and U.S. Patent No. 4,631,211 to Houghten, *et al.* (1986)).

Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art (see, for instance, Sutcliffe, *et al.*, *supra*; Wilson, *et al.*, *supra*; Chow, M., *et al.*, *Proc. Natl. Acad. Sci. USA* **82**:910-914; and Bittle, F. J., *et al.*, *J. Gen. Virol.* **66**:2347-2354 (1985)). Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art (see, for instance, Geysen, *et al.*, *supra*). Further still, U.S. Patent No. 5,194,392, issued to Geysen, describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a

particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092, issued to Geysen, describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971, issued to Houghten and colleagues, on Peralkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

As one of skill in the art will appreciate, follistatin-3 polypeptides of the present invention and the epitope-bearing fragments thereof described above can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life *in vivo*. This has been shown, e.g., for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (EP A 394,827; Traunecker, *et al., Nature* **331**:84-86 (1988)). Fusion proteins that have a disulfide-linked dimeric structure due to the IgG part can also be more efficient in binding and neutralizing other molecules than the monomeric follistatin-3 protein or protein fragment alone (Fountoulakis, *et al., J. Biochem.* **270**:3958-3964 (1995)).

Follistatin-3 protein-specific antibodies for use in the present invention can be raised against the intact follistatin-3 protein or an antigenic polypeptide fragment thereof, which may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier.

Antibodies

Further polypeptides of the invention relate to antibodies and T-cell antigen receptors (TCR) which immunospecifically bind a polypeptide, polypeptide fragment,